Molecular Karyotyping by Array CGH

Linking gene dosage alterations to disease phenotypes

Lisenka ELM Vissers

Molecular karyotyping by array CGH - Linking gene dosage alterations to disease phenotypes
The studies presented in this thesis were performed at the Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

ISBN/EAN 978-90-9022012-3

Cover photo: Zorginbeeld.nl/Frank Muller
Print: Gildeprint Drukkerijen BV, Enschede

Molecular Karyotyping by Array CGH

Linking gene dosage alterations to disease phenotypes

Een wetenschappelijke proeve op het gebied van de Medische Wetenschappen

Proefschrift

ter verkrijging van de graad doctor
aan de Radboud Universiteit Nijmegen,
op gezag van de Rector Magnificus prof. mr. S.C.J.J. Kortmann
volgens besluit van het College van Decanen
in het openbaar te verdedigen op

donderdag 13 september 2007
om 13.30 uur precies

door

Lisenka Elisabeth Louisa Maria Vissers

geboren op 27 augustus 1980
te Vught

Promotor:
Prof. dr. A. Geurts van Kessel

Copromotor:
Dr. ir. J.A. Veltman

Manuscriptcommissie:
Prof. dr. J. Schalkwijk (voorzitter)
Prof. dr. B.C.J. Hamel
Prof. dr. C. Wijmenga (Rijksuniversiteit Groningen)

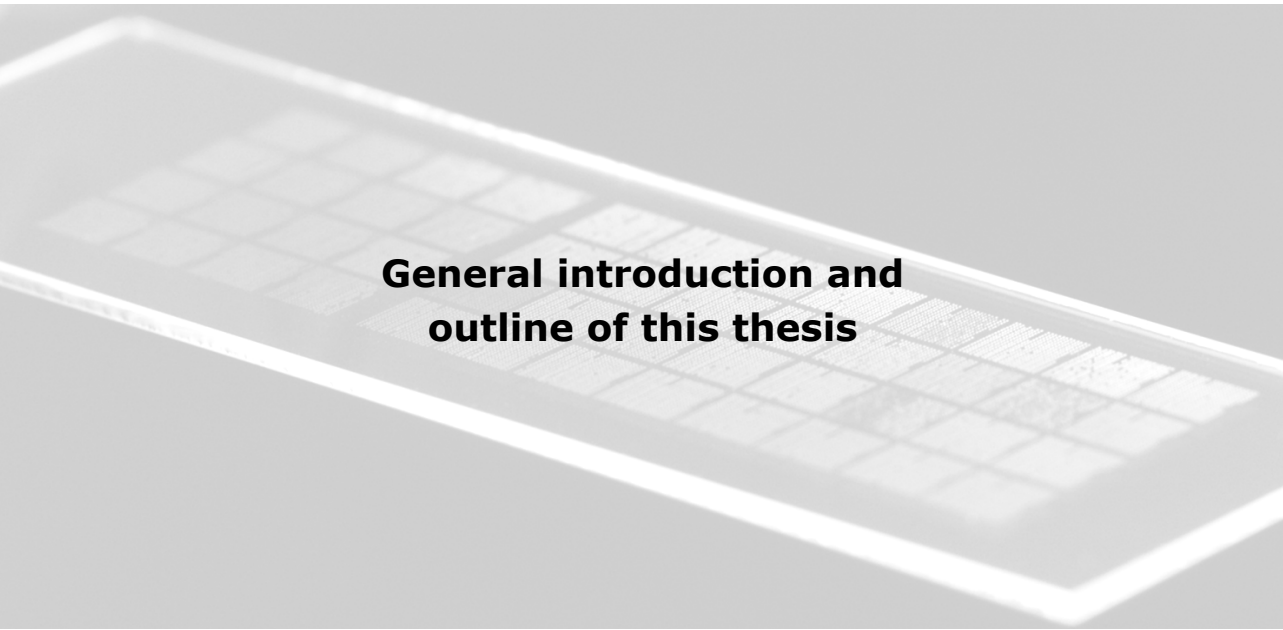
Contents

Abbreviations

A	Adenine
Array CGH	Microarray-based Comparative Genomic Hybridization
BAC	Bacterial Artificial Chromosome
bp	base pair
BERA	Brainstem Evoked Response Audiometry
C	Cytosine
CCD	Charged Coupled Device
CCR	Complex Chromosome Rearrangement
cDNA	complementair Desoxyribo Nucleic Acid
CGH	Comparative Genomic Hybridization
CHD	chromodomain helicase DNA-binding
CMT1A	Charcot-Martie Tooth Disease Type 1A
CNS	Conserved Non-coding Sequence
CNV	Copy Number Variation
CT	Computer Tomography
DAPI	4,6-diamidino-2-phenylindole
DECIPHER	DatabasE of Chromosome Imbalance and Phenotype in Humans Using Ensembl Resources
DNA	Desoxyribo Nucleic Acid
DMSO	Dimethyl sulfoxide
DOP	Degenerate oligonucleotide-primed
ECARUCA	European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations
EMG	Electromyography
FISH	Fluorescent *In Situ* hybridization
FSHD1A	FacioScapuloHumeral Dystrophy 1A
G	Guanine
HMM	Hidden Markov Model
HNPP	Hereditary Neuropathy with liability to Pressure Palsies
IQ	Intelligence Quotient
kb	kilobase (Thousand base pairs)
LCR	Low-copy repeat
LGTC	Leiden Genome Technology Center
LINEs	Long Interspersed Nuclear Elements
LOWESS	LOcally WEighted Scatterplot Smoothing
Mb	Megabase (Million base pairs)
MDLS	Miller-Dieker Lissencephaly Syndrome
M-FISH	Multicolour Fluorescent *in situ* hybridization
MLPA	Multiplex Ligation-dependent Probe Amplification
MR	Mental Retardation
MRI	Magnetic Resonance Imaging
mRNA	messenger RiboNucleic Acid
NAHR	NonAllelic Homologous Recombination
NHEJ	Non-Homologous End Joining
OMIM	Online Mendelian Inheritance in Man

ORF	Open Reading Frame
PAC	P1-derived Artifical Chromosome
PCR	Polymerase Chain Reaction
PWM	Periventricular White Matter
PWS	Prader Willi Syndrome
RNA	RiboNucleic Acid
ROMA	Representative Oligonucleotide Microarray Analysis
RPCI	Roswell Park Cancer Institute
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SD	Standard Deviation
SINEs	Short Interspersed Nuclear Elements
SKY	Spectral karyotyping
SMA	Spinal Muscular Dystrophy
SMS	Smith Magenis Syndrome
SNP	Single Nucleotide Polymorphism
SSEP	Somatosensory Sensory Evoked Potentials
T	Thymine
TRPS	TrichoRhinoPhalangeal Syndrome
UCSC	University of California, Santa Cruz
WV	Wide Ventricles
ZonMW	Nederlandse organisatie voor gezondheidsonderzoek en zorginnovatie

Chapter 1



General introduction and outline of this thesis

Based on:

Lisenka ELM Vissers, Joris A Veltman, Ad Geurts van Kessel & Han Brunner
Identification of disease genes by whole genome CGH arrays.
Human Molecular Genetics 14 Spec No.2: R215-223, 2005

Joris A Veltman & Lisenka ELM Vissers
Molecular karyotyping by means of array CGH: Linking gene dosage alterations to disease phenotypes.
In: DNA Microarrays, edt. U Nuber, BIOS Scientific Publishers, UK, 2005

1.1 The human karyotype

During the first 40 years of the previous century, scientists tended to dismiss the possibility that DNA could carry the genetic information in chromosomes, partly because nucleic acids were wrongly believed to contain only a simple repeating nucleotide sequence consisting of four different nucleotides, i.e., adenine (A), cytosine (C), thymine (T) and guanine (G). However, we now know that DNA is a long, linear molecule that contains millions of these four nucleotides arranged in an irregular but non-random order, providing a blueprint for the formation and maintenance of an organism. This blueprint represents the genome of an organism.

The DNA molecule constitutes a double helix, which is coiled and folded into a chromosome structure. Chromosomes can be visualized during cell division using light microscopy. In 1956, cytogenetics, the study of chromosome sets and their genetic properties, revealed that the human cell contains 46 chromosomes, consisting of 22 pairs of autosomes, and a single pair of sex chromosomes (Tjio & Levan 1956). The subsequent development of chromosome banding techniques such as Giemsa-, Quinacrine Hydrochloride- and Reverse banding, has enabled a very precise recognition of individual chromosomes (Caspersson et al. 1968; Caspersson et al. 1970; Rooney 2001; Trask 2002; Smeets 2004). Ordering the chromosomes according to their size and position of the centromere, in conjunction with their banding pattern, is referred to as karyotyping. A normal female karyotype is designated as 46,XX whereas a normal male karyotype is designated as 46,XY.

Chromosomal landmarks, such as size, position of the centromere and banding pattern, are not only useful for the identification of normal chromosomes, but also for the characterization of structural abnormalities in them which, next to numerical abnormalities, may seriously hamper the processes of normal growth and development.

1.2 Microscopically visible chromosome abnormalities

Numerical abnormalities result from loss or gain of individual chromosomes and/or sets of chromosomes. Loss of one chromosome copy gives rise to a monosomy for that particular chromosome, whereas gain of a chromosome copy results in a trisomy. A monosomy is almost always incompatible with embryonic development, with the exception of monosomy X which causes Turner's syndrome. Trisomies for the autosomes 13, 18 or 21 are compatible with survival to pregnancy term, but trisomy 13 and trisomy 18 result in severe conditions that convey a poor prognosis. Consequently, most affected newborns die during the first few weeks of life. Individuals with trisomy 21, better known as Down's syndrome, have a better prognosis and their current life expectancy is 50 to 60 years (Mueller & Young 2001).

Structural abnormalities result from chromosome breakage and subsequent joining in an altered configuration. This altered configuration can lead to loss or gain of genetic material, in which case the abnormality is unbalanced. Alternatively, all genetic material may be retained. In this latter case the abnormality is balanced. Structural chromosome abnormalities include (i) deletions, (ii) duplications, (iii) isochromosomes, (iv) ring chromosomes, (v) inversions and (vi) translocations (**Figure 1.1**). By definition, deletions, isochromosomes, duplications and ring chromosomes are unbalanced in nature.

Translocations and inversions can either be balanced or unbalanced. The clinical consequences of unbalanced structural rearrangements depend on the size of the rearrangement and the genetic information affected. For example, a deletion of a single dosage-sensitive gene can cause a clinically recognizable syndrome (paragraph 1.6). However, a balanced rearrangement may give rise to a similar clinical phenotype if it disrupts the function of such an important gene and/ or results in the formation of a new fusion gene.

Both numerical and structural chromosome abnormalities can be present in all cells of the body or, alternatively, in only a subset of them. Constitutional abnormalities that are present in all cells most likely result from a gamete containing the abnormality, whereas abnormalities that are present in only a subset of cells may have been acquired during a later stage of development. As a consequence, acquired abnormalities are present in a mosaic configuration, introducing a second cell line with a different chromosomal constitution. The clinical consequences of acquired chromosome abnormalities depend on the developmental stage and the cell type in which these abnormalities occur. For example, if the abnormality occurs during early embryogenesis affecting a progenitor cell, this may lead to a severe congenital defect. Alternatively, if the abnormality occurs later in life, this somatic event may lead to the aberrant proliferation of a cell that normally would no longer replicate. The clone of mutant cells generated in this way may eventually progress towards cancer.

Gross chromosomal rearrangements can be detected by karyotyping. However, this approach has its limitations. First, karyotyping requires mitotic cells in order to obtain chromosomes in their optimal configuration for visualization (i.e. metaphase chromosomes). Second, karyotyping has a limited resolution, i.e., structural chromosome rearrangements need to be of sufficient size in order to be detectable (involving at least 5-10 million base pairs, Mb). Third, the detection of abnormalities may be difficult, in particular in cases where only a small percentage of cells contains the abnormality.

1.3 Detection of submicroscopic chromosome rearrangements

The resolution of chromosome analysis has greatly benefited from the introduction of fluorescent *in situ* hybridization (FISH) techniques (van Prooijen-Knegt et al. 1982). FISH allows the detection of rearrangements beyond the detection limit of conventional karyotyping. The technology relies on the unique ability of single stranded fluorescently labeled DNA, known as probe, to anneal to its complementary sequence in chromosomes. Next, the location of the annealed (or hybridized) probe can be visualized by use of a fluorescent microscope. Depending on the application, different types of FISH probes can be used, including telomere-specific probes, whole-chromosome painting probes and/or locus-specific probes. The latter probe type can be used to e.g. specifically analyse chromosomal regions based on clinical suspicion.

The resolution of chromosome analysis was further enhanced by the possibility to simultaneously hybridize a multitude of probes. By doing so, the staining of all human chromosomes in different colours has become possible, and various techniques to do so are currently available such as multicolour FISH (M-FISH) and spectral karyotyping (SKY) (Liehr et al. 2004). These latter FISH-based technologies have the advantage of studying

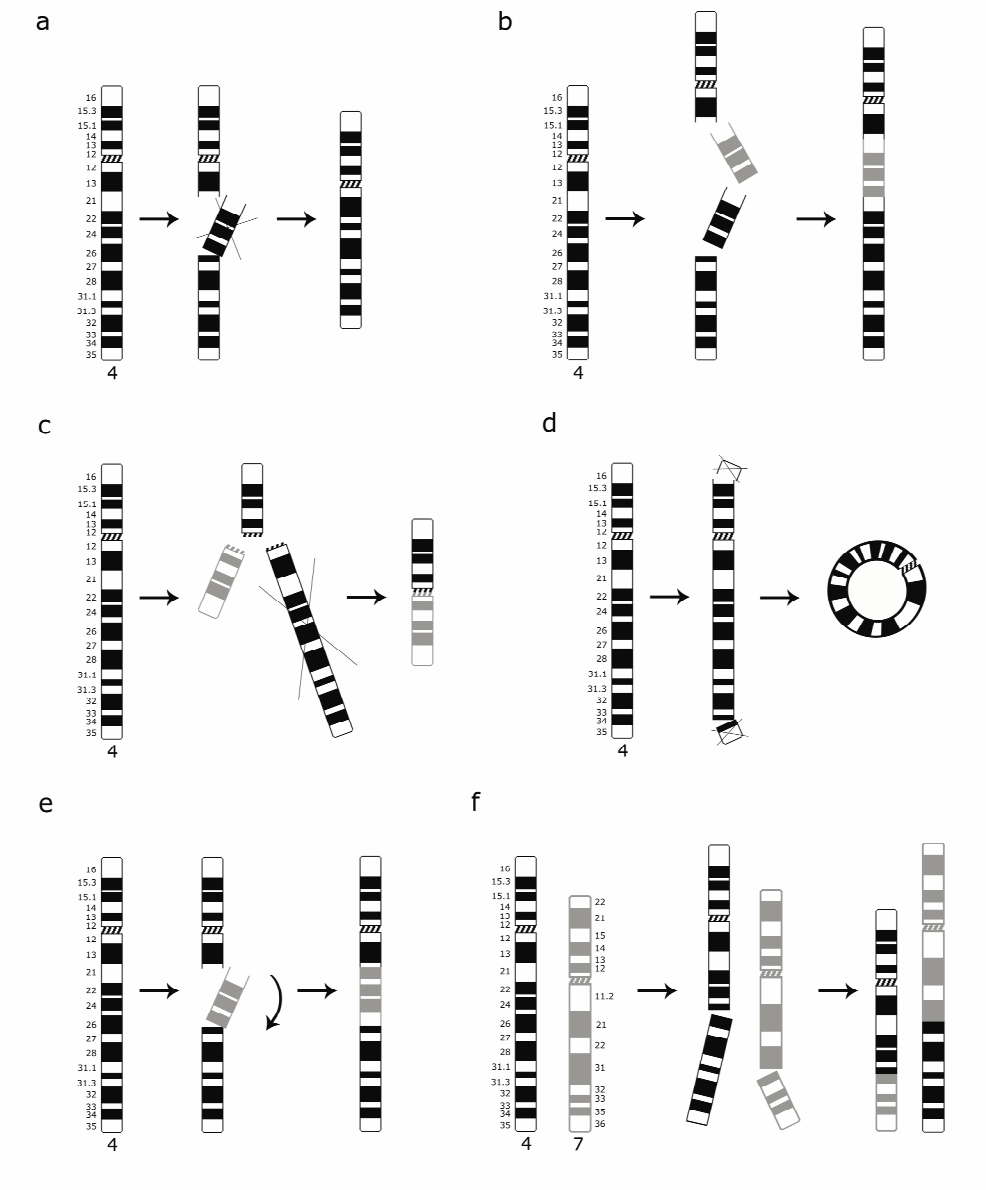


Figure 1.1
Schematic representation of structural rearrangements. (**a**) Deletion. Loss of genetic material. (**b**) Duplication. Insertion of genetic material (grey). (**c**) Isochromosome of the p arm. Loss of one chromosome arm with duplication of the other arm (duplication in grey). (**d**) Ring chromosome. Joining of two 'sticky' chromosome ends caused by deletion of genetic material on both chromosome arms. (**e**) Inversion. Reversion of genetic material on a single chromosome. (**f**) Translocation. Transfer of genetic material from one chromosome to another chromosome. In this case a translocation of the chromosomes 4 (in black) and 7 (in grey). Part of the 4q arm is exchanged with material originating from the 7q arm.

the constitution of all human chromosomes in a single experiment, and thus, are particularly useful for the analysis of (complex) chromosome rearrangements without prior knowledge of the chromosomal regions involved.

In subsequent years, these genome-wide FISH technologies were further modified, resulting in the introduction of a method that has been denoted comparative genomic hybridization (CGH) (Kallioniemi et al. 1992; Lichter et al. 2000). CGH is based on the comparison of two genomic DNA populations, one derived from a test (patient) sample, and one derived from a normal reference sample. Equal amounts of DNA are differentially labeled and simultaneously hybridized onto normal human metaphase chromosomes, thus competing for the same targets on the chromosomes. Variation in fluorescence intensities of test and reference DNA along each chromosome target reveals the genomic locations of chromosome rearrangements in the test DNA. The advantage of CGH over M-FISH or SKY is its independence of actively dividing cells from the test sample as a source for metaphase spreads. As such, CGH can be performed on virtually all samples from which DNA can be extracted. CGH has proven particularly useful in cancer research. In general, tumour samples are difficult to culture and harvest for preparing metaphase spreads. In addition, these spreads are difficult to analyse by conventional karyotyping because of the abundance and complexity of the rearrangements present. The resolution of CGH, however, still depends on the resolution of the target metaphase chromosomes, i.e., it remains difficult to detect rearrangements below the level of 5-10 Mb (Forozan et al. 1997).

1.4 Clinical consequence of submicroscopic chromosome rearrangements

Over the last decades various FISH studies have revealed that submicroscopic subtelomeric rearrangements account for approximately 6% of all previously unexplained cases of mental retardation (MR) (Flint et al. 1995; Knight et al. 1999; Knight & Flint 2000; de Vries et al. 2003). Similarly, it was found that interstitial submicroscopic chromosome rearrangements account for a vast proportion of contiguous gene syndromes (paragraph 1.6; Osborne et al. 2001; Shaikh et al. 2001). However, in cases where the clinical phenotype has not previously been associated with a known genomic rearrangement, there is no *a priori* knowledge of the region to be tested and, hence, FISH is no longer the method of choice. In order to detect such submicroscopic rearrangements on a genome-wide scale, novel technologies were needed that combine the resolution of targeted FISH technologies with the genome-wide approach of CGH. One such technology is array-based comparative genomic hybridization (array CGH).

1.5 Array-based comparative genomic hybridization

Through the development of novel technologies such as array CGH the resolving power of conventional chromosome analysis techniques has increased from the megabase to the kilobase level (Solinas-Toldo et al. 1997; Pinkel et al. 1998). Tools that have facilitated the development of these technologies include (i) genome-wide clone resources integrated into the finished human genome sequence, (ii) high-throughput microarray platforms, and (iii) optimized CGH protocols and data analysis systems. Together, these microarray-based

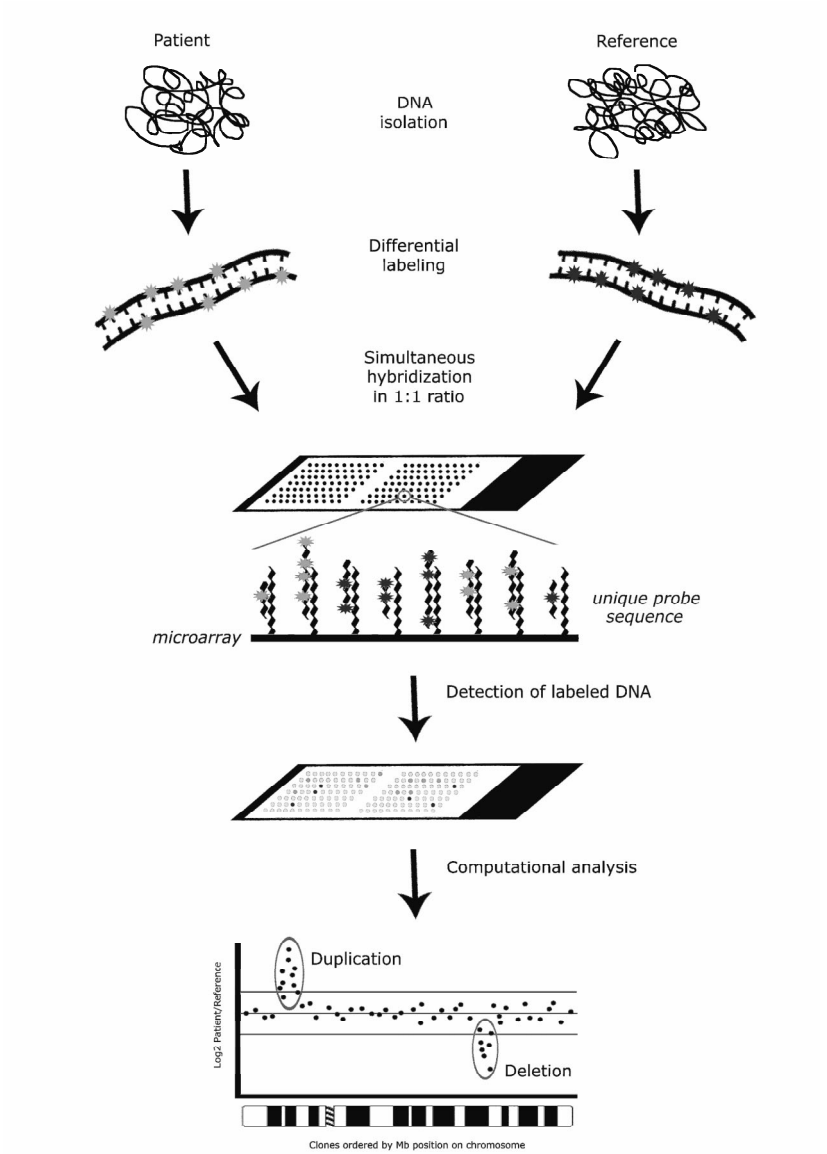


Figure 1.2
Schematic overview of the array CGH procedure. Genomic DNA samples from a test (patient; left) and reference (normal control; right) are differentially labeled with different fluorochromes, usually Cy3 and Cy5 (for green and red, respectively). The two DNA samples are mixed in equal amounts and hybridized to the microarray, onto which large-insert clone DNAs (e.g. BAC clones) have been robotically spotted as targets. Subsequent computer imaging assesses the relative fluorescence levels of each labeled DNA for each array target. Clones to which equal amounts of patient DNA and reference DNA have been hybridized, will appear in yellow, clones deleted in the patient DNA compared to the reference DNA will appear in red, and clones that are duplicated in the patient DNA compared to the reference DNA will appear in green. The ratio between the test and reference DNA for each clone on the array can be plotted to visualize chromosome imbalances, marked by a deviation from the normal $\log_2$ T/R ratio.

technological developments have accumulated into a so-called 'molecular karyotyping' approach that allows for the sensitive and specific detection of submicroscopic single copy number changes throughout the entire human genome. Array CGH builds upon conventional CGH procedures (paragraph 1.3), in such a way that the target metaphase spreads are replaced by genomic fragments with known physical locations in a microarray format. In comparison with conventional CGH, the microarray format provides a higher resolution, a higher dynamic range and a better possibility for automation. In addition, it allows for direct linking of (submicroscopic) chromosome rearrangements to known genomic sequences and, thus, to genes which may be involved in the disease under investigation.

The array CGH procedure that employs BAC clones as targets includes the following steps. First, the large DNA insert BAC clones are amplified by either degenerate oligonucleotide-primed (DOP) PCR (Telenius et al. 1992) or ligation-mediated PCR (Klein et al. 1999) in order to obtain sufficient quantities needed for spotting. Next, these PCR products are spotted onto glass slides coated with substrates such as aminosilane using microarray robots equipped with high-precision printing pins. Depending on the amount of clones to be spotted and the space available on the microarray slide, clones can either be spotted once per array or in replicate. Test (patient) and reference (normal control) DNAs are usually labeled with either Cy3- or Cy5-dUTP using random priming and, subsequently, co-hybridized onto the microarray in a solution containing an excess of Cot1-DNA to block repetitive sequences. During co-hybridization, test and reference DNA compete for the same targets on the microarray. Subsequently, the hybridized DNAs are detected by the different fluorochromes using scanning equipment with either a confocal laser or a charge coupled device (CCD) camera-based reader, followed by spot identification. The presence and location of chromosomal imbalances can be detected and quantified by calculating the ratio of the test-over-reference fluorescent intensities along the targets on the microarray. These T/R ratios show which chromosomal regions in the test genome are overrepresented (duplicated) or underrespresented (deleted) relative to the reference genome. As the ratio profiles can be ordered based on the physical location of the target BAC clones, an overview of the (submicroscopic) chromosomal imbalances throughout the entire genome can be generated, thereby providing a 'molecular karyotype' (**Figure 1.2**).

The increase in data obtained through high-density microarrays requires standardized storage systems as well as thorough statistical tools, similar to those required for microarray-based gene expression profiling (Quackenbush 2002; Stoeckert et al. 2002). Owing to the complicated process of producing and hybridizing spotted microarrays, a certain degree of systematic variation does exist in the data produced. Normalization of microarray data is used to eliminate such systematic variation and, therefore, represents an important pre-processing step in the analysis of almost all microarray data. One of the most frequently applied normalization procedures in microarray-based expression studies is the locally weighted scatterplot smoothing (LOWESS). After this data normalization, automated statistical procedures are required for the reliable detection of genomic copy number changes. One such algorithm uses the Hidden Markov Model (HMM) (Rabiner 1989) which is not only suited for distinguishing genuine copy number changes from random microarray noise, but also for precisely localizing the start- and end-points of each copy number alteration. Independent validation of chromosome rearrangements may be obtained either via FISH using probe sequences targeted to the region involved, or via

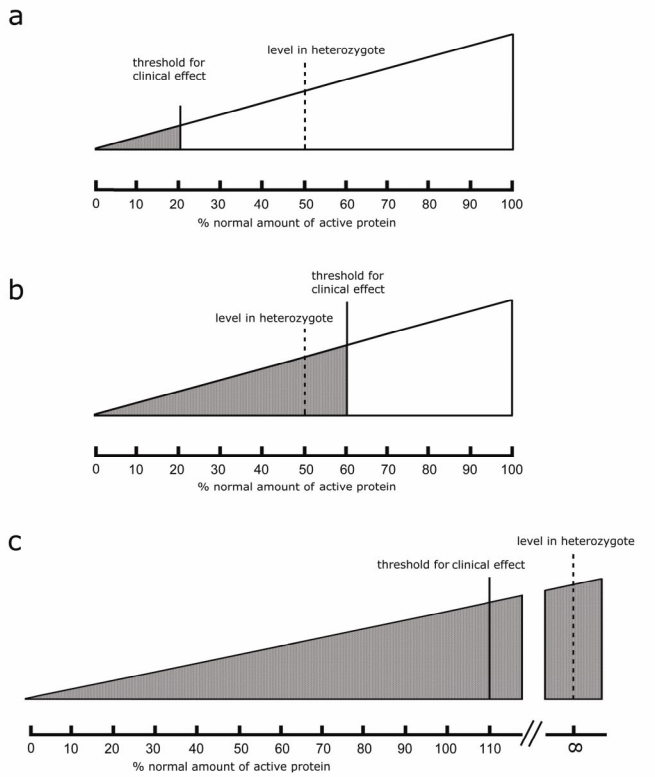


Figure 1.3
Putative effects of mutations that alter the quantity of a gene product (protein). The solid vertical lines mark the thresholds for clinical effects, whereas the dotted vertical lines mark the levels of protein present. (**a**) In case of a heterozygous mutation, the clinical effect may become apparent when the protein level is below 20% of normal. (**b**) In case of haploinsuffiency, the clinical effect may become apparent when the protein level is below 60% of normal. (**c**) In case of a gain-of-function mutation, the clinical effect may become apparent when the level of product is above 100% of normal.

multiplex ligation-dependent probe amplification (MLPA), a semi-quantitative PCR assay using primers specifically designed to the region of interest (Schouten et al. 2002). This comprehensive array CGH technique, once established and validated, allows the high-throughput genome-wide screening for chromosomal imbalances with a resolution limited only by the size and distribution of the target BAC clones used.

1.6 Monogenic disease, contiguous gene syndromes and genomic disorders

In single gene disorders, or monogenic diseases, the phenotypic spectrum observed can be attributed to the malfunctioning of a single gene. Such malfunctioning can be achieved by physical deletion or duplication of (parts of) a genomic copy of the gene or by more

subtle intragenic mutations. The ultimate effect of a deletion or mutation that e.g. leads to a premature stop in the reading frame of the affected gene is haploinsufficiency, a state by which a decrease in the level of the corresponding protein gives rise to the phenotype. Such genes are dosage-sensitive (**Figure 1.3**). In reverse, duplications or gain-of-function mutations create proteins that exhibit an increase in constitutive activity, even in the absence of a physiological activator, or that create insensitivity to negative regulators (**Figure 1.3**). To date, over 10,000 single gene disorders are known and listed in the database of Online Mendelian Inheritance in Man (OMIM).

In contrast to single gene disorders, it has been shown that several conditions, including mental retardation and additional congenital/developmental abnormalities, may be due to submicroscopic chromosome rearrangements encompassing several genes. In 1986, Schmickel (1986) coined the term contiguous gene syndrome for these disorders. Since the introduction of this term many alternatives have been suggested including 'microdeletion/ microduplication syndrome' and 'segmental aneusomy syndrome'. All these terms intend to imply that the phenotype of the disorder results from an inappropriate dosage of more than one critical gene located within the genomic region affected, i.e., individual genes located in such genomic regions contribute to distinct clinical features of the syndrome. It has, therefore, been suggested that the extent of the chromosomal region involved in each case would correlate with the ultimate phenotype, and that individual clinical features might be inherited in isolation (Budarf & Emanuel 1997).

Interestingly, for many 'microdeletion/microduplication syndromes', the submicroscopic deletion or duplication identified appeared to be of virtual identical size in all patients. In-depth molecular analyses have shown that local genomic architectural features, such as low-copy repeat (LCR) structures, play an important role in the formation of these recurrent microdeletions and microduplications. LCRs, also termed segmental duplications or duplicons, encompass apparently normal stretches of genomic DNA and often contain genes. LCRs are present in more than one copy in the genome and are defined by a ≥ 1 kb size and a >90% sequence identity (Bailey et al. 2001; Eichler 2001; Bailey et al. 2002). LCRs can mediate recurrent DNA rearrangements such as deletions, duplications and inversions through chromosome or chromatid misalignment followed by nonallelic homologous recombination (NAHR) (**Figure 1.4**). In fact, the majority of the currently known 'microdeletion/duplication syndromes' result from NAHR between large (usually >10 kb), highly identical (>95%) LCRs and are, therefore, also termed genomic disorders (Lupski 1998; Stankiewicz & Lupski 2002; Lupski & Stankiewicz 2006).

1.7 Disease-gene identification in monogenic disease and contiguous gene syndromes

Microscopically visible and submicroscopic chromosome rearrangements have significantly contributed to the delineation of single gene disorders. For instance, translocations of Xp21 suggested for the first time that the Duchenne muscular dystrophy gene might map to this chromosomal region (Jacobs et al. 1981). While *de novo* translocations have been most widely used for the mapping and identification of disease genes, deletions have also been instrumental for the positional cloning of a variety of disease genes, including those for familial adenomatous polyposis (Herrera et al. 1986), retinoblastoma (Lele et al. 1963), WAGR syndrome (Riccardi et al. 1978), and a number of contiguous gene

Table 1.1: Monogenic diseases with frequent occurrence of deletions or duplications >50 kb.

Syndrome	Location	Dosage-sensitive gene	Frequency of deletion/ duplication	Reference
Sotos syndrome	5q35	*NSD1*	6-49 %[a]	Kurotaki et al. 2003
Neurofibromatosis Type 1	17q11.2	*NF1*	7 %	Cnossen et al. 1997
Alagille syndrome	20p12	*JAG1*	7 %	Krantz et al. 1997
Rubinstein-Taybi syndrome	16p13.3	*CREBBP*	10 %	Petrij et al. 2000
Congenital 21-alpha hydroxylase deficiency	6p21.3	*CYP21A2*	28 %	Olney et al. 2002
Transient neonatal diabetes	6q24		36 %	Temple et al. 1996
Cystinosis syndrome	17p13	*CTNS*	44 %	Shotelersuk et al. 1998
Pelizaeus-Merzbacher	Xq22	*PLP1*	62 %	Mimault et al. 1999
Smith-Magenis syndrome	17p11.2	*RAI1*	>90 %	Juyal et al. 1996

[a] :mutation/deletion detection dependent on ethnic background

syndromes (Schmickel 1986). One particularly successful application of systematic deletion analysis has been the identification of a number of genes for holoprosencephaly, including *SHH, ZIC2, SIX3* and *TGIF* (Munke 1989; Roessler et al. 1996; Brown et al. 1998; Wallis et al. 1999; Gripp et al. 2000).

As outlined above, it is becoming increasingly clear that many 'microdeletion syndromes' can largely or completely be attributed to the phenotypic effects of haploinsufficiencies for single genes (paragraph 1.6). Pertinent examples are the *RAI1* gene in Smith-Magenis syndrome (Slager et al. 2003), the *UBE3A* gene in Angelman syndrome (Kishino et al. 1997), and the *TBX1* gene in deletion 22q11 syndrome (Lindsay et al. 2001). For the *LIS1* gene in Miller-Dieker syndrome, however, the situation is more complex. While deletion of this gene is responsible for lissencephaly (Dobyns et al. 1993), it was found that the concomitant deletion of the *14-3-3 epsilon* gene also contributes to this brain phenotype (Cardoso et al. 2003; Toyo-Oka et al. 2003). The reason that such clinical conditions are usually caused by microdeletions and rarely by intragenic mutations reflects their local genomic architecture rather than the intrinsic features of the causative gene(s) itself (Stankiewicz & Lupski 2002; paragraph 1.6). In fact, the only real requirement for a candidate microdeletion syndrome gene is that it should be dosage-sensitive. In case of microduplications, the effect of having a complete extra copy of a gene may result in a phenotype that is not mirrored by other mutations in this gene. For example, *PMP22* gene duplications result in Charcot-Marie-Tooth Type 1A whereas point mutations in this gene may lead to hereditary liability to pressure palsies (Valentijn et al. 1992; Chance et al. 1993). However, this does not hold for all cases, since both duplications and deletions of the *PLP* gene are common causes of Pelizaeus-Merzbacher disease (Woodward & Malcolm 1999). Also, deletions and duplications of the *SOX3* gene appear to yield a similar phenotype, i.e., infundibular hypoplasia and hypopituitarism (Woods et al. 2005).

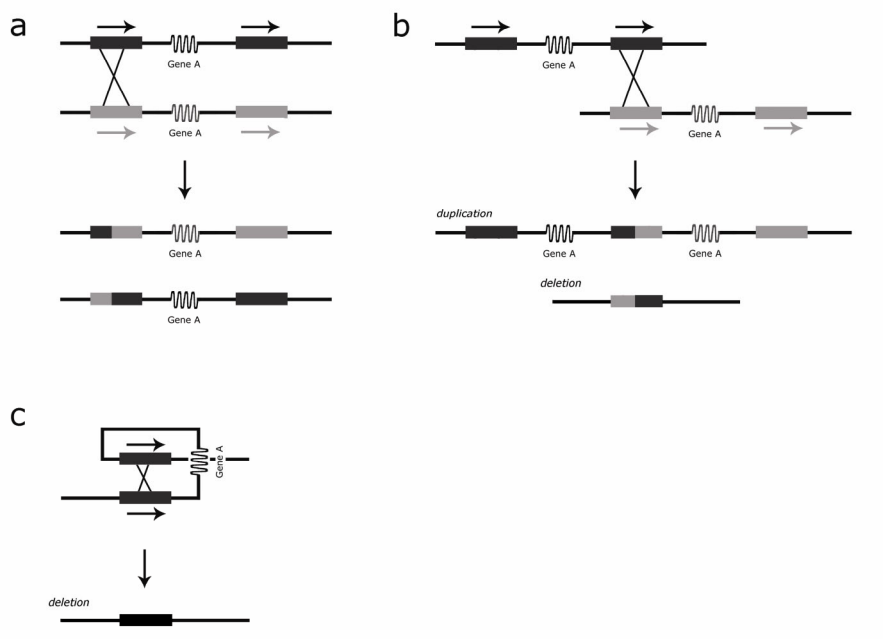


Figure 1.4
Model for nonallelic homologous recombination (NAHR) involving region-specific low-copy repeats (gray blocks). (**a**) Normal situation in which the appropriate LCRs align and recombine. (**b**) Interchromosomal NAHR of LCRs on two misaligned homologous chromosomes, resulting in a (submicroscopic) duplication and a reciprocal deletion of the intervening DNA segment. (**c**) Intrachromosomal NAHR of LCRs forming an intrachromatidal loop, resulting in the excision of the intervening segment. Arrows indicate the relative orientation of LCRs to each other.

The frequencies at which microdeletions or microduplications are encountered in monogenic diseases differ markedly. For example, there are monogenic diseases that are mostly caused by gene mutations and rarely by deletions or duplications, such as von Recklinghausen Neurofibromatosis, Rubinstein-Taybi syndrome and Alagille syndrome (**Table 1.1**). In other monogenic diseases, however, large deletions or duplications involving a dosage-sensitive gene are responsible for the majority of the cases (**Table 1.1**). A more complex situation is encountered in Sotos syndrome. This syndrome is caused predominantly by heterozygous *NSD1* point mutations in the Caucasian population (Douglas et al. 2003), whereas microdeletions containing the *NSD1* gene prevail in the Japanese population (Kurotaki et al. 2003). This difference in mutation spectrum may reflect differences in genomic architecture between Japanese and Caucasians. Thus, microdeletions and microduplications occur at various frequencies in many monogenic diseases with a known genetic cause, and the difference between a microdeletion syndrome with rare mutations and a single gene disorder with occasional large deletions may be gradual rather than absolute. The availability of novel, highly sensitive methods, such as array CGH (paragraph 1.5), may further enhance the possibilities for a straightforward mapping of the genes underlying these disorders.

1.8 Aim and outline of this thesis

The work described in this thesis was initiated in September 2002, five years after the first hybridization of patient-derived DNA to nucleic acids from different genomic loci spotted on a microarray was reported (array CGH; Solinas-Toldo et al. 1997). In those five years, the application of array CGH remained mainly restricted to cancer research (Pinkel et al. 1998; Pollack et al. 1999; Albertson et al. 2000; Snijders et al. 2001; Wilhelm et al. 2002). Based on this experience, it was anticipated that array CGH-based genome-wide copy number screening might also be applicable to the detection of submicroscopic chromosome rearrangements in patients with mental retardation and/or malformation syndromes. Within this context, it was of particular relevance that the potency of this technique for the identification of chromosome rearrangements without any prior knowledge of the genomic regions involved would reach far beyond that of the existing chromosome analysis techniques. In addition, the integration of large collections of BAC clones into genome browsers like those from Ensembl and UCSC would allow for a direct inspection of candidate genes affected by the copy number imbalances detected which, in turn, should be of major help in explaining the clinical phenotypes. Therefore, the general aim of this thesis was to explore the feasibilities of the array CGH technology for its application in both clinical genetic and genome research.

Specifically, we aimed at the following objectives:

I. The generation of a microarray for genome-wide copy number assessment and optimization of protocols for the detection of single copy chromosome imbalances.

The detection of submicroscopic chromosome deletions and duplications might enhance the mapping of genes underlying monogenic diseases and, simultaneously, facilitate the identification of novel contiguous gene syndromes. In **Chapter 2** we describe the generation of a microarray containing ~3,500 FISH-verified BAC clones selected to cover the human genome with an average spacing of one clone per megabase. After optimisation, we used this array for the identification of submicroscopic chromosome rearrangements in 20 patients with unexplained mental retardation. In this series, three microdeletions and two microduplications were uncovered.

II. The application of array CGH for resolving the genetic causes of well-characterized clinical syndromes

For several well-recognized clinical syndromes, the genetic cause remains elusive. The identification of recurrent microdeletions and microduplications may resolve the genetic cause for such syndromes. In addition, microdeletions encompassing causative genes have been reported to occur at low frequencies in single gene disorders. In **Chapter 3** we show that high-resolution genome-wide screening by array CGH can be used as an effective new approach for this purpose through the identification of the gene underlying CHARGE syndrome.

III. The application of array CGH for the identification of novel 'microdeletion/ microduplication syndromes'.

The identification of overlapping submicroscopic deletions or duplications in patients who display similar phenotypic characteristics, might reflect new microdeletion syndromes or

genomic disorders. In **Chapter 4** we describe the identification of three overlapping interstitial submicroscopic deletions on 17q21.31 in patients with a previously unknown clinically recognizable syndrome. Interestingly, the microdeletions are directly related to the local genomic architecture of 17q21.31, thereby classifying this new syndrome as a genuine genomic disorder.

IV. The assessment of large-scale DNA copy number variations within the human genome.

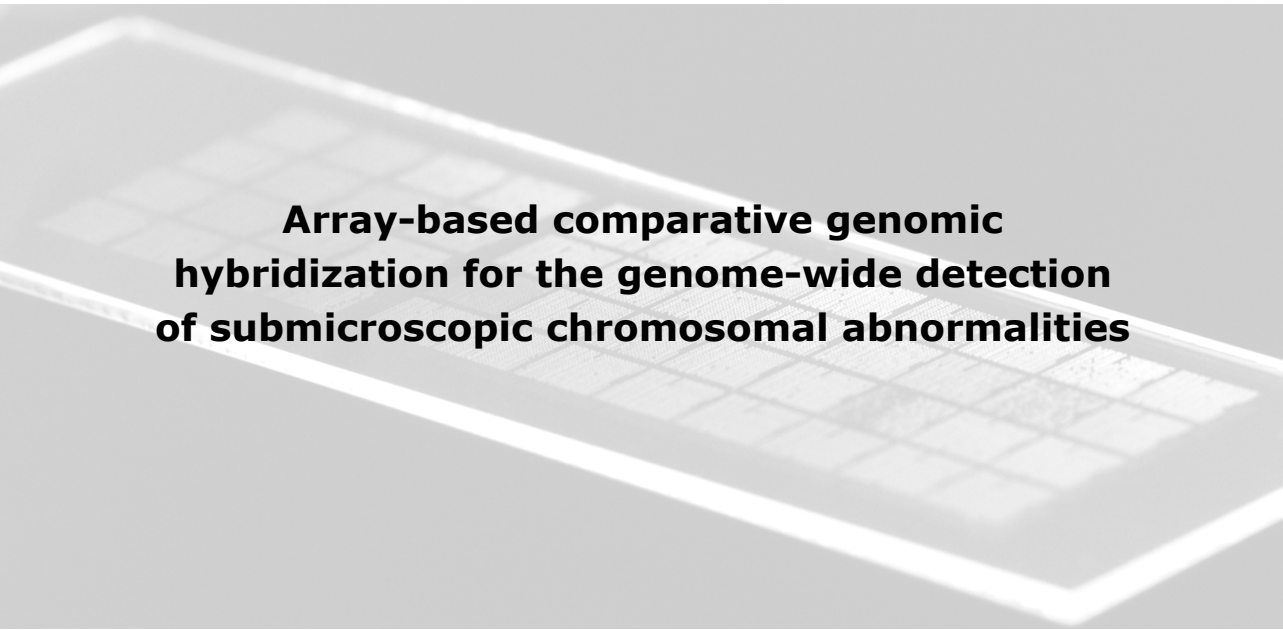
By using array CGH, a new class of genomic variations has been uncovered (Iafrate et al. 2004; Sebat et al. 2004). These variations may encompass genomic segments ranging in size from ~100 kb to ~2.5 Mb, and are commonly referred to as copy number variations (CNVs). The full nature and extent of CNV and its frequency in different ethnic populations is, however, largely unknown. In **Chapter 5** we describe a detailed analysis of 12 CNV loci in >300 healthy individuals of five different ethnic populations. From this analysis we conclude that several of these CNV loci differ between populations, not only in frequencies between populations, but also in absolute DNA copy numbers within populations. Our data suggest that insight into absolute DNA copy numbers for loci exhibiting CNV is a prerequisite for assessing their putative role in health and disease.

V. The role of genomic architecture in mediating non-recurrent submicroscopic chromosome rearrangements.

Nonallelic homologous recombination between region-specific genomic architectural features is a major mechanism underlying the occurrence of recurrent chromosomal rearrangements associated with many genomic disorders. However, the question whether such architectural features also mediate non-recurrent chromosomal rearrangements has largely remained elusive. In **Chapter 6** we describe a detailed molecular analysis of two individuals with non-recurrent complex chromosome rearrangements (CCR) involving the short arm of chromosome 17. In total, we identified and characterized 17 breakpoints by high-resolution array CGH and dual-color FISH analysis. Of these 17 breakpoints, four were located within low copy repeats (LCRs), suggesting that these genomic architectural features may indeed mediate the formation of these non-recurrent chromosomal rearrangements.

In **Chapter 7** the objectives explored in the preceding chapters are discussed and an outline for future research is presented. Finally, a summary of this thesis is provided in **Chapter 8**.

Chapter 2



Array-based comparative genomic hybridization for the genome-wide detection of submicroscopic chromosomal abnormalities

Lisenka ELM Vissers, Bert BA de Vries, Kazutoyo Osoegawa, Irene M Janssen, Ton Feuth, Chik On Choy, Huub Straatman, Walter van der Vliet, Erik HLPG Huys, Anke van Rijk, Dominique Smeets, Conny MA van Ravenswaaij-Arts, Nine V Knoers, Ineke van der Burgt, Pieter J de Jong, Han G Brunner, Ad Geurts van Kessel, Eric FPM Schoenmakers & Joris A Veltman

American Journal of Human Genetics 73: 1261–1270, 2003

Abstract

Microdeletions and microduplications, not visible by routine chromosome analysis, are a major cause of human malformation and mental retardation. Novel high-resolution, whole-genome technologies can improve the diagnostic detection rate of these small chromosomal abnormalities. Array-based comparative genomic hybridization allows such a high-resolution screening by hybridizing differentially labeled test and reference DNAs to arrays consisting of thousands of genomic clones. In this study, we tested the diagnostic capacity of this technology using ~3,500 fluorescent *in situ* hybridization–verified clones selected to cover the genome with an average of 1 clone per megabase (Mb). The sensitivity and specificity of the technology were tested in normal-versus-normal control experiments and through the screening of patients with known microdeletion syndromes. Subsequently, a series of 20 cytogenetically normal patients with mental retardation and dysmorphisms suggestive of a chromosomal abnormality were analyzed. In this series, three microdeletions and two microduplications were identified and validated. Two of these genomic changes were identified also in one of the parents, indicating that these are large-scale genomic polymorphisms. Deletions and duplications as small as 1 Mb could be reliably detected by our approach. The percentage of false-positive results was reduced to a minimum by use of a dye-swap-replicate analysis, all but eliminating the need for laborious validation experiments and facilitating implementation in a routine diagnostic setting. This high- resolution assay will facilitate the identification of novel genes involved in human mental retardation and/or malformation syndromes and will provide insight into the flexibility and plasticity of the human genome.

Introduction

Mental retardation, with or without additional malformations, occurs in 2–3% of the general population. Although a considerable number of cases can be explained by the presence of gross chromosomal abnormalities or other factors, such as metabolic and/or neurological anomalies, the etiology of mental retardation remains unexplained for ~50% of patients (Anderson et al. 1996; de Vries et al. 1997). Submicroscopic, subtelomeric chromosome rearrangements contribute significantly to mental retardation and malformation, comprising up to 6% of the previously unexplained cases (Flint et al. 1995; Knight et al. 1999; Biesecker 2002; de Vries et al. 2003). These findings underscore the potential importance of submicroscopic chromosomal anomalies as a major cause of human mental retardation and malformation. To routinely detect these changes in a diagnostic setting, an efficient and robust technology is needed that screens the entire genome for copy number abnormalities with a resolution beyond the level of a light microscope (5–10 Mb). Array-based comparative genomic hybridization (array CGH) technology measures submicroscopic DNA copy number changes and allows the simultaneous high-resolution mapping of these changes onto the genome sequence (Solinas-Toldo et al. 1997; Pinkel et al. 1998; Snijders et al. 2001). We previously developed an array-based subtelomeric assay that screens all human subtelomeric regions in a single hybridization reaction (Veltman et al. 2002). Here we report the construction and application of a genome-wide microarray for the identification of known and novel microdeletions and duplications in patients with mental retardation and malformations.

Patients and Methods

Patients

Genomic DNAs, isolated from blood lymphocytes of four cytogenetically normal, healthy individuals (two males and two females), were used for array validation and as normal reference DNAs. Additional genomic DNAs were isolated from three patients with FISH-verified known microdeletion syndromes (Prader-Willi syndrome [PWS; MIM 176270] on chromosome 15q11-15q12, Smith-Magenis syndrome [SMS; MIM 182290] on chromosome 17p11.2, and Trichorhinophalangeal syndrome [TRPS; MIM 190350] on chromosome 8q23.1–q24.11), as well as from 20 patients with mental retardation and additional dysmorphisms of unknown etiology. The latter patients were all seen by a clinical geneticist and had undergone extensive diagnostic work-up, including routine chromosome analysis without a diagnosis. They all had a phenotype suggestive of a chromosomal abnormality, and all scored three points or higher on the checklist developed by de Vries et al. (2001). Genomic DNAs from patients and controls were isolated and purified using a QIAamp kit (Qiagen), according to the instructions of the manufacturer.

Array-Based comparative genomic hybridization

Clone set

A total of 3,569 well characterized, colony-purified, and FISH-verified BAC clones were used for array construction. Most of the BACs were derived from the RPCI-11 BAC library used as the main intermediate substrate for the sequencing and mapping of the human genome (Osoegawa et al. 2001). The set includes ~3,200 clones selected through a collaboration of the Children's Hospital Oakland Research Institute, BACPAC Resources Center, and several other groups to cover the genome with a 1-Mb resolution (Cheung et al. 2001). Information on this clone set and its availability can be obtained at the BACPAC Resources Center Web site. Additional clones were added to the array, resulting in an even higher-resolution coverage of genomic regions known to be involved in human malformation and mental retardation, including the subtelomeric regions of all human chromosomes (77 clones) (Knight et al. 2000) and regions associated with known microdeletion syndromes (30 clones). Finally, chromosome 12 and chromosome 18 were covered with a higher density through the addition of clones used in previous studies (Veltman et al. 2003a; Zafarana et al. 2003).

Array preparation

Genomic target DNAs were isolated from 12-ml bacterial cultures using Qiagen R.E.A.L. Prep 96 BioRobot kits on a Qiagen BioRobot 9600 (Qiagen), following the instructions of the manufacturer. Degenerate oligonucleotide-primed (DOP) PCR was performed on isolated DNA from all clones, essentially as described elsewhere (Telenius et al. 1992), with minor modifications (Veltman et al. 2002). Taq2000 (Stratagene) was used as a thermostable polymerase. DOP-PCR products were dissolved at a concentration of 1 μg/μl in a 50% DMSO solution and robotically spotted in triplicate onto CMT-GAPS coated glass slides (Corning, UltraGaps) using an OmniGrid 100 arrayer (Genomic Solutions). The array consisted of 48 subgrids, and replicates were printed in different subgrids across the array.

Labeling and hybridization

Labeling and hybridization were performed essentially as described elsewhere (Veltman et al. 2002). In brief, genomic DNA was labelled by random priming with Cy3-dUTP or Cy5-dUTP (Amersham Biosciences). Test and reference samples were mixed with 120 μg Cot-1

DNA (Roche), coprecipitated, and resuspended in 130 µl of a hybridization solution containing 50% formamide, 10% dextran sulfate, 2 × SSC, 4% SDS, and 10 µg/µl yeast tRNA (Invitrogen). After denaturation of probe and target DNA, hybridization and posthybridization washing procedures were performed using a GeneTAC Hybridization Station (Genomic Solutions), according to the manufacturer's instructions. In brief, an 18-h hybridization with active circulation of the probe was performed, followed by five posthybridization wash cycles in 50% formamide/2 × SSC at 45ºC and five wash cycles in phosphate-buffered saline at 20ºC. Slides were dried by centrifugation after a brief wash in water.

Image analysis and processing

Slides were scanned and imaged on an Affymetrix 428 scanner (Affymetrix) using the Affymetrix 428 scanner software package (version 1.0). The acquired microarray images were analyzed using GenePix Pro 4.0 (Axon Instruments), as described elsewhere (Veltman et al. 2002). For all further analyses, the median of the pixel intensities minus the median local background was used for every spot on the array (Cy3 and Cy5, calculated separately). Data normalization was performed in the software package SAS version 8.0 (SAS Institute) for each array subgrid, by applying Lowess curve fitting with a smoothing factor of 0.1 to predict the $\log_2$-transformed test-over-reference (T/R) value on the basis of the average logarithmic fluorescent intensities (Cleveland 1979). This smoothing factor was shown to result in the lowest percentage of false-positive results while not increasing the amount of false-negative results in the validation experiments. A consequence of this smoothing procedure is that the ratios of the clones with a copy number gain or loss are closer to the normal range of $\log_2$ ratios than in normalization procedures without this smoothing.

Quality control

Clones with an SD of the triplicates >0.3 were excluded in individual experiments, as well as clones with fewer than two replicates remaining after this analysis. Excluded from all experiments were 63 clones that did not show reliable hybridization results in at least four of the five normal-versus-normal control experiments. Clones that mapped to the sex chromosomes (n=163) were not analyzed in detail. The array contains a final set of 3,343 autosomal clones with a coverage of at least 1 clone per Mb. From this final set, 3% of the clones, on average, were excluded per experiment on the basis of the quality criteria. Thresholds for copy number gain and loss were determined by examining the results of the control experiments and of previously published work and were set at $\log_2$ T/R values of 0.3 and -0.3, respectively. Experiments were excluded when >5% of the clones showed intensity ratios outside of these regions. Of the 40 experiments performed in this study, 5 experiments did not meet these quality criteria. These experiments were successfully repeated. The final data set is available as a downloadable electronic supplement via the online version of this article.

Analysis of replicate experiments

In this study, we performed a dye-swap experiment for each case (patient or control). For statistical analysis of these two experiments, we developed a two-dimensional assay in the software package SAS version 8.0 (SAS Institute) in which reference regions were calculated containing 99.999% of the data points (*P=.99999*), assuming that the pairs of normalized ratios follow a bivariate normal distribution (**Figure 2.1b** and **Figure 2.1d**). Under the assumption of no deleted or duplicated regions, the number of data points

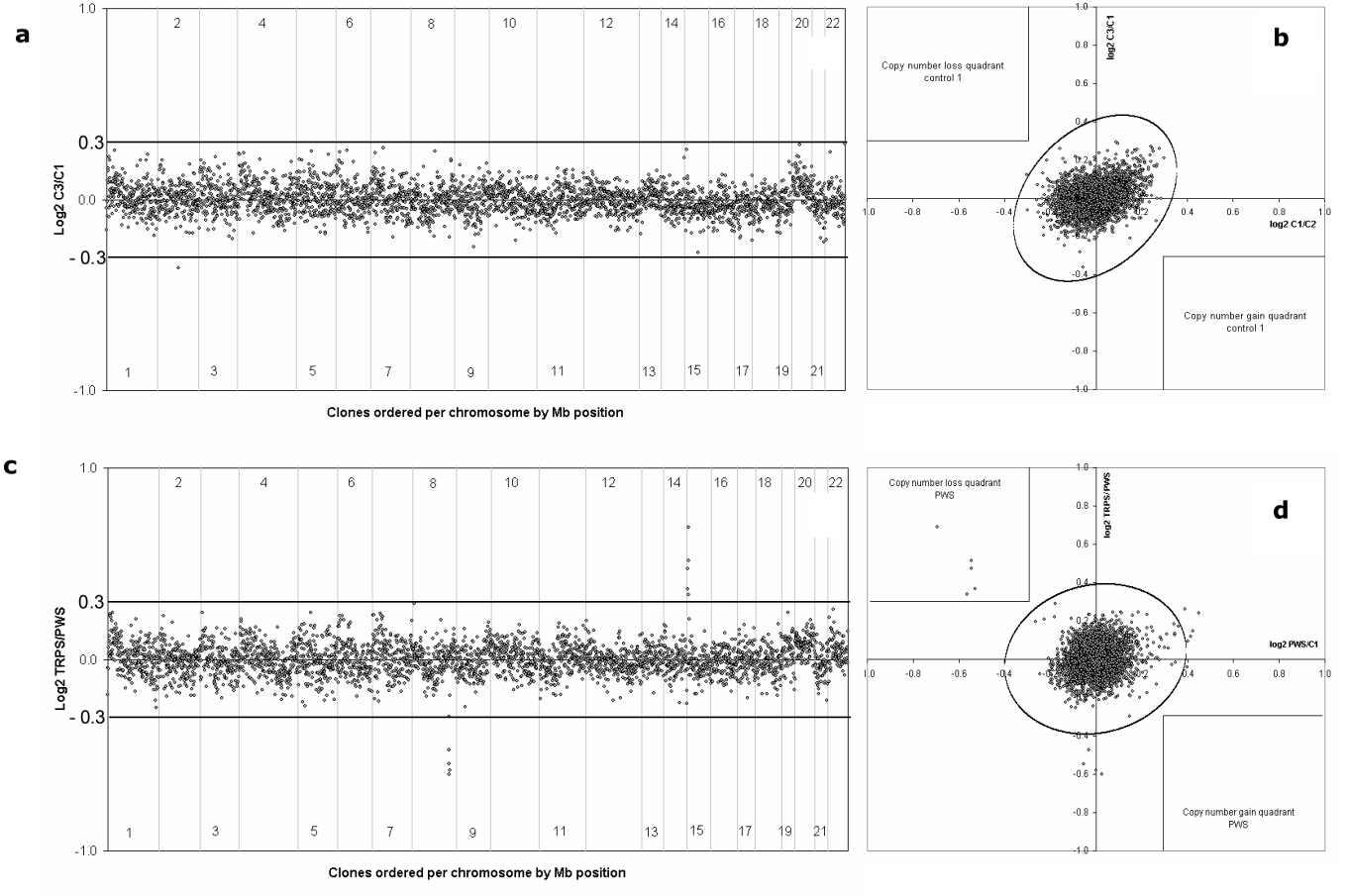
outside the resulting ellipse is expected to be 1/100,000 × the number of clones on the array— in our case, 1/100,000 × 3,343=0.03. Clones represented by data points outside this reference region in the scatterplot are candidates for a microduplication or deletion event. However, since a dye-swap experiment was performed for each case, the data points also have to be located in the correct quadrant of the scatterplot (i.e., a positive sign for experiment 1 [patient 1 vs. control 1] and a negative sign for experiment 2 [control 1 vs. patient 1] indicates a potentially duplicated clone, whereas a deleted clone shows opposite signs in both experiments). The *a priori* thresholds for copy number gain ($\log_2$ T/R value 0.3) or loss ($\log_2$ T/R value -0.3) are therefore integrated into the scatterplot to indicate the candidate clones for microdeletion or duplication events.

FISH validation experiments

FISH validation experiments were performed on metaphase spreads prepared from patient-derived lymphoblast cell lines using routine procedures. Probe labeling, slide preparation, and hybridization were carried out essentially as described elsewhere (de Bruijn et al. 2001). A Zeiss epifluorescence microscope, equipped with appropriate filters, was used for visual examination of the slides. Digital images were captured using a high-performance cooled CCD camera (Photometrics) coupled to a Macintosh Quadra 950 computer. The Image FISH software package (Intergen) was used for analysis of the FISH images. Inverted images of DAPI-stained slides were used for chromosome identification.

Figure 2.1

Array CGH genomic profiles of validation experiments. Arrays contained 3,343 human autosomal clones (indicated by small circles representing the mean $\log_2$-transformed and Lowess-normalized T/R intensity ratios), ordered in **a** and **c** from 1pter to 22qter on the basis of the physical mapping positions obtained from the November 2002 freeze of the UCSC genome browser. In panels **a** and **c**, chromosome boundaries are indicated by vertical lines. Panel **a** shows the result of a normal-versus-normal hybridization (control 3 vs. control 1). Nearly all clones fall within the *a priori* thresholds for copy number gain ($\log_2$ T/R value 0.3) and copy number loss ($\log_2$ T/R value -0.3) indicated by the horizontal lines. One clone on chromosome 2 shows an intensity ratio outside these thresholds and might represent a false-positive result. Panel **b** shows the result of the combined analysis of the two hybridizations performed with control 1 (*x* axis: control 1 vs. control 2; *y* axis: control 3 vs. control 1). The ellipse represents the border of the reference regions containing 99.999% of the data points; the thresholds for copy number gain and loss are also integrated into this figure (see the "Patients and Methods" section for details). As can be seen, there is only one clone outside the reference region; however, this clone does not pass the thresholds for copy number loss in both experiments and can therefore be discarded from further analyses. The clone on chromosome 2 that fell outside the threshold for copy number loss in panel **a** is clearly within the normal reference region and can therefore also be discarded for further analyses. Panel **c** shows the result of the hybridization of DNA from a patient with TRPS against DNA from a patient with PWS. A total of four clones, spanning 2.7 Mb of genomic sequence on 8q23.3-q24.11, showed $\log_2$ TRPS-over-PWS intensity ratios below the threshold for copy number loss, confirming the presence of a deletion of this genomic region in the TRPS patient. In addition, five clones, spanning 2.9 Mb of sequence on 15q11.2, show $\log_2$ intensity ratios above the (reverse) threshold for copy number gain, indicating a deletion of this genomic region in the PWS patient. No clones outside these target genomic regions show potential false-positive results. The combined results of two experiments involving the PWS patient are shown in panel **d**. The five target clones on 15q11.2 are reproducibly deleted in both experiments and fall outside the bivariate normal distribution reference region (*P=.99999*) and within the copy number loss quadrant indicated in the upper left quadrant.


a
b
c
d
Log2 C3/C1
Clones ordered per chromosome by Mb position
Copy number loss quadrant control 1
Copy number gain quadrant control 1
log2 C3/C1
log2 C1/C2
Log2 TRPS/PWS
Copy number loss quadrant PWS
Copy number gain quadrant PWS
log2 TRPS/PWS
log2 PWS/C1

Results

Validation experiments

To test the specificity and sensitivity of the whole genome BAC array, we performed a series of five normal-versus-normal control hybridizations using four normal healthy blood donors (including a dye-swap experiment for each control). **Figure 2.1a** shows a representative genomic profile resulting from such an experiment. Nearly all clones show $\log_2$ intensity ratios in between the *a priori* thresholds for copy number gain (0.3) or loss (-0.3). In the five normal-versus-normal experiments, an average of six clones (0.18%) passed these thresholds. Although very low, this level of background noise would still require a substantial number of FISH experiments to distinguish true microdeletions and microduplications from false-positive results. Therefore, the combination of two experiments (with dye swap) was analyzed for each control case, using stringent criteria for the presence of copy number gain or loss (see the "Patients and Methods section" and **Figure 2.1b**). The added value of combining data of two separate (T/R) experiments was clearly shown by the fact that the number of false positives was reduced to zero in all four cases tested.

Next, we tested the sensitivity of the technology by hybridizing DNA from three patients with known, FISH-confirmed microdeletion syndromes (i.e., one patient with PWS, one patient with SMS, and one patient with TRPS) to the genome-wide array. Similar to the normal-versus-normal experiments, DNA samples were hybridized not only against each other but also against one of the normal controls. **Figure 2.1c** shows the result of a hybridization of DNA from the patient with TRPS against DNA from the patient with PWS. A total of four clones, representing 2.7 Mb of consensus genomic sequence on 8q23.3-q24.11, showed TRPS-over-PWS intensity ratios below the threshold for copy number loss, thus confirming the presence of a deletion of this genomic region in the TRPS patient. In addition, five clones, spanning 2.9 Mb of sequence on 15q11.2, showed $\log_2$ intensity ratios above the (reverse) threshold for copy number gain, indicating a deletion of this genomic interval encompassing the genes *SNRPN* and *UBE3A* in the patient with PWS. The combined results of two experiments involving the patient with PWS are shown in **Figure 2.1d**. The five target clones on 15q11.2 are reproducibly deleted in both experiments and fall outside the bivariate normal distribution reference region (*P=.99999*) and within the upper left quadrant, indicating copy number loss. Detailed analysis of the DNA from these two patients and the patient with SMS (containing a deletion of two genomic clones spanning a 1.5-Mb region on 17p11.2) showed that the target microdeletion region could be readily identified in each individual experiment with the target clones present on the array. Individual hybridizations showed an average of seven clones (0.21%) with $\log_2$ intensity ratios outside the thresholds for copy number gain or loss, very similar to the percentage of false positives obtained in the normal-versus-normal control experiments described above. Statistical analysis of the duplicate experiments for each case failed to reproduce any of these aberrant ratios, strongly indicating that these were indeed false-positive results.

In conclusion, the sensitivity to detect submicroscopic (1.5–2.9 Mb) deletions was reproducibly validated, and the specificity of the technology was assured by performing two hybridizations for each case and applying a stringent statistical analysis.

Table 2.1: Phenotypes and Detected Microdeletions/-duplications

Patient	Sex	Age (years)	Degree of MR[a]	Phenotype[b]	Array CGH result
1	Male	2	Severe	Microcephaly, facial dysmorphism, pulmonary arterial and valve stenosis	del(7)(q11.21q11.23), 17 clones, 8 Mb, de novo
2	Female	12	Severe	Microcephaly, facial dysmorphism, short stature	del(2)(q22.3q23.2), 3 clones, 2 Mb, de novo
3	Female	23	Moderate	Facial dysmorphism, short stature, ventricular septum defect	del(1)(p21p21), 1 clone, <1 Mb, also detected in father
4	Male	2	Severe	Microcephaly, facial dysmorphism	dup(2)(q21.2q21.2), 1 clone, <1 Mb, also detected in father
5	Male	19	Mild	Facial dysmorphism, polydactyly, polycystic kidney (familial)	dup(6)(q25.3q26), 1 clone, <1 Mb, parents not available

[a]: MR=mental retardation. [b]: Detailed clinical information is given in the text.

Detection of novel deletions and duplications in patients with unexplained mental retardation and dysmorphisms

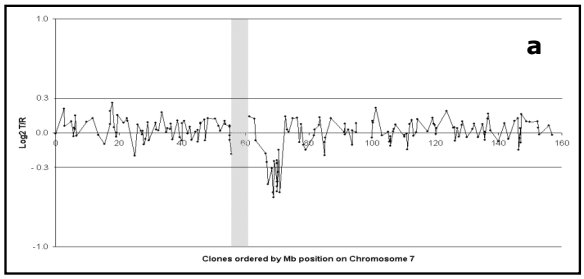
We selected a test series of 20 patients with mental retardation of unknown etiology to investigate whether the high-resolution, genome-wide, microarray-based, copy number screening would allow the identification of small genomic changes not detectable by routine karyotyping. In concordance with the validation experiments, each case was hybridized once against a normal control sample and once against another patient with a different clinical phenotype. This procedure was chosen to minimize the costs and, at the same time, the risks of hybridizing cases with identical genetic abnormalities against each other, which would result in masking of the abnormality. In 7 of the 20 patients, copy number alterations were reproducibly detected by array CGH on our genome-wide microarray. In five of these cases, the results could be confirmed by FISH on metaphase spreads of the patients. These included three microdeletions and two microduplications (**Table 2.1; Figure 2.2**). *De novo* occurrence was checked by investigating DNA samples from the parents. As expected, the microduplications proved difficult to validate by FISH. We therefore repeated the array CGH procedure in the two cases with microduplications on a small, high density array containing only the regions of interest and confirmed the presence of a microduplication in both cases (data not shown). The largest deletion identified in patient 1 was verified by FISH (**Figure 2.2a** and **Figure 2.2f**) and targeted 17 clones on the array, spanning a total region of 8.6 Mb on 7q11.21-q11.23. The karyotype of this case was re-examined because of the relatively large size of this deletion, but no abnormalities were identified. It is interesting that this genomic region

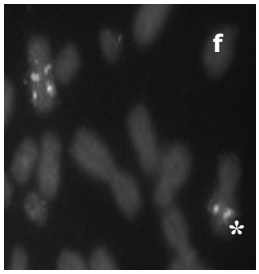
contains the complete common deletion segment for the Williams-Beuren syndrome (Bayes et al. 2003). The other deletions and duplications were considerably smaller. In patient 2, the deletion on 2q22.3-q23.2 was encompassed by three clones mapping in a 2-Mb genomic interval, whereas, in the other three cases, only one clone was involved, indicating the presence of abnormalities <1Mb in size. In one patient (patient 3) with a deletion on 1p21, the deletion was identified by FISH in the father of the patient as well, indicating that this might be a novel genomic polymorphism. Similarly, the duplication of a single clone on 2q21.2 in patient 4 was identified by FISH in the father. Unfortunately, the parents of patient 5, whose DNA contained a duplication of a single clone on 6q25.3-q26, were not available for checking *de novo* occurrence.

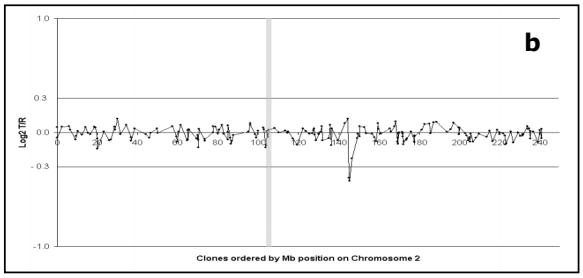
Discussion

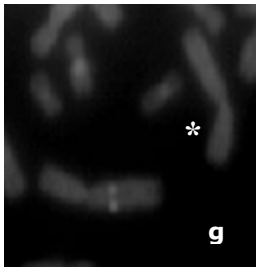
Array CGH provides a high spatial genomic resolution and allows a fully automated evaluation of thousands of genomic loci. Previous applications have been mainly directed at genomic abnormalities in cancer (Snijders et al. 2001; Veltman et al. 2003b; Wessendorf et al. 2003). In this study, we demonstrate the application of array CGH in detecting known and novel submicroscopic abnormalities. The specificity and sensitivity of this approach was tested and validated in cytogenetically normal and healthy individuals, as well as in patients with known microdeletion syndromes. Deletions and duplications were detected reliably in a single overnight hybridization experiment without a priori knowledge of the genomic region involved. Individual experiments were performed with a low level of false-positive results that was reduced further by performing a replicate dye-swap experiment followed by a thorough statistical analysis. From our data, we conclude that such replicate experiments are essential for implementation in a diagnostic setting, since they considerably reduce the need for laborious confirmation experiments while greatly improving the validity of the results. Application of this approach in a pilot study of 20 patients with unexplained mental retardation and additional malformations resulted in

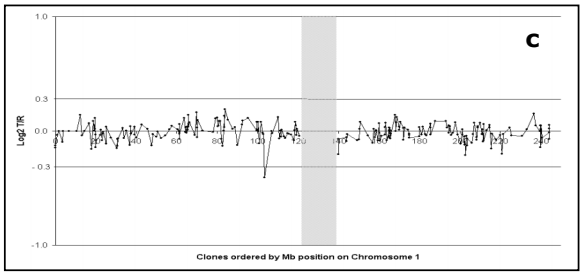
Figure 2.2
Detailed genomic profiles and FISH validation of copy number abnormalities identified in five cases with unexplained mental retardation. Panels **a–e** represent individual profiles of the affected chromosomes for each case, with clones ordered, for each chromosome, from pter to qter, on the basis of the physical mapping positions obtained from the November 2002 freeze of the UCSC genome browser. The centromeric region is indicated by a vertical gray dash, the thresholds for copy number gain ($\log_2$ T/R value 0.3) and copy number loss ($\log_2$ T/R value -0.3) are indicated by horizontal lines. Panels **f-j** represent the FISH validation using (one of) the target clone(s) identified by array CGH. Affected chromosomes are indicated by an asterisk (*). Panel **a** shows the deletion on 7q11 in patient 1, with 14 clones in this region showing an average $\log_2$ intensity ratio of -0.5. FISH validation of this case is shown in the adjacent panel **f**, in which one of the deleted clones on 7q11 is shown in red and an undeleted control probe is shown in green. Panel **b** shows the microdeletion on 2q22 in patient 2 with a total of three clones crossing the threshold for copy number loss, with FISH validation in the adjacent panel **g**. Deletion of a single clone on 1p21 is shown in panel **c** for patient 3; this clone was confirmed by FISH to be deleted not only in the patient (panel **h**) but also in the father of the patient. Copy number gain detected in a single clone is shown in panels **d** and **e** for patients 4 and 5, with FISH confirmation in panels **i** and **j**.


a
Log2 TIR
1.0
0.3
0.0
- 0.3
-1.0
0
20
40
60
80
100
120
140
160
Clones ordered by Mb position on Chromosome 7

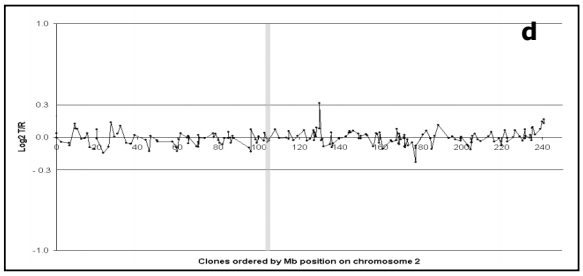

f
*


b
Log2 TIR
1.0
0.3
0.0
- 0.3
-1.0
0
20
40
60
80
100
120
140
160
180
200
220
240
Clones ordered by Mb position on Chromosome 2

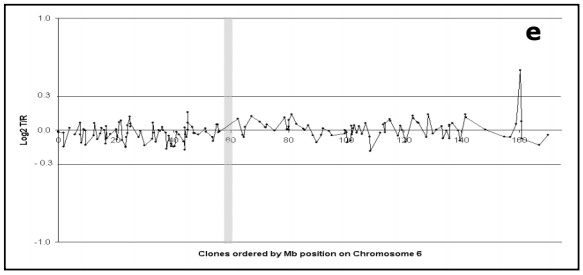

*
g


c
Log2 TIR
1.0
0.3
0.0
- 0.3
-1.0
0
20
40
60
80
100
120
140
160
180
200
220
240
Clones ordered by Mb position on Chromosome 1


*
h


d
Log2 TIR
1.0
0.3
0.0
- 0.3
-1.0
0
20
40
60
80
100
120
140
160
180
200
220
240
Clones ordered by Mb position on chromosome 2


*
i


e
Log2 TIR
1.0
0.3
0.0
- 0.3
-1.0
0
20
40
60
80
100
120
140
160
Clones ordered by Mb position on Chromosome 6


*
j

the detection of five copy number abnormalities, three deletions, and two duplications, all beyond the microscopic resolution (~10 Mb). The patient with the largest deletion (8.6 Mb on 7q11.21-q11.23) was a 2-year-old boy (patient 1) with severe mental retardation, plagio- and microcephaly, postnatal growth retardation, facial dysmorphism (downward-slanting palpebral fissures, periorbital fullness, epicanthus and telecanthus, broad mouth with full lips, and sagging cheeks), short neck, unilateral simian crease, and a peripheral and valvular pulmonary stenosis. The deletion completely overlaps with the common 1.6 Mb region deleted in patients with Williams syndrome (Bayes et al. 2003). Although this boy had some facial features fitting the diagnosis of Williams syndrome, his clinical presentation was more severe than is commonly observed in patients with Williams syndrome, most notably the severe retardation. Moreover, a pulmonary valve stenosis is less often observed in this syndrome (Eronen et al. 2002). It seems likely that the severity of the phenotype is related to the large size of the deletion, a finding that has also been observed in four cases with even larger deletions, three of which also showed characteristics typical of the Williams syndrome (Valentine & Sergovich 1977; Frydman et al. 1986; Mizugishi et al. 1998; Wu et al. 1999).

A 2-Mb deletion on 2q22.3-q23.2 was identified in a 12-year-old girl (patient 2) who had severe mental retardation, short stature (height 3 SDs below normal), microcephaly (head circumference 2.5 SDs below normal), obesitas, facial dysmorphism (coarse facies, upward- slanting palpebral fissures, hypotelorism, abnormally shaped ears, high nasal bridge, small, carp-shaped mouth with downward-turned corners, narrow, flat palate, and broad chin), and long, narrow hands with short digiti V. The deletion was just distal to a more common deletion on 2q22 and did not include the *SIP1* gene, which is associated with Mowat-Wilson syndrome (Mowat et al. 2003). The clinical presentation of this girl differs also from the phenotype seen in this relatively new syndrome. The deletion is *de novo*, so haploinsufficiency of one or more genes within the deleted region could be causative for the phenotype. So far, no obvious candidate gene appears to be present in this genomic region.

In a 23-year-old female (patient 3) with moderate mental retardation, autism, short stature, minor facial dysmorphism (upward-slanting palpebral fissures, deepset eyes, short philtrum), short broad feet, a small ventricular septum defect, and childhood absences, we detected a deletion of a single clone on 1p21. This deletion was present in her healthy father, as well. Although familial occurrence of this small deletion (<1 Mb) does not rule out a causative role, it may very well be a novel genomic polymorphism. It is known that similar polymorphisms (either deletions or duplications) without any clinical significance are present in the genome, but, at present, they are underrecognized. Examples of known normal genomic variations include a 2.5-Mb duplication of 8p23.1 (Barber et al. 1998; Engelen et al. 2000) and a number of subtelomeric polymorphisms (Ballif et al. 2000; Linardopoulou et al. 2001; Der-Sarkissian et al. 2002).

It is interesting that two microduplications were also detected using this comparative microarray technology. One duplication on 2q21.2 (<1 Mb) occurred in a 2-year-old boy (patient 4) with severe mental retardation, microcephaly, and facial dysmorphism (metopic ridge, synophrys, arched eyebrows, long eyelashes, upwardslanting palpebral fissures, low-set, posteriorly rotated malformed ears, and high nasal bridge). This duplication was detected also in his father, indicating the presence of another genomic

polymorphism. The other duplication was found to be present on 6q25.3-q26 (<1 Mb) in a 19-year-old male (patient 5) with mild mental retardation, postaxial polydactyly of hands and feet, facial dysmorphism (upward-slanting palpebral fissures, high, narrow nasal bridge, short philtrum, retrognathia, and irregular teething), and medullary polycystic kidneys. The kidney abnormality was also present in his two mentally normal sisters. It is interesting to note that, in a series of 36 patients with a larger duplication, including the 6q25.3-q26 region, four cases have been reported with a polydactyly (Schinzel 2001). Unfortunately, the parents of this patient were unavailable for checking de novo occurrence. However, a common copy-number variation affecting a single BAC at 6q26 was recently reported by Albertson and Pinkel (2003). In this publication, the observed copy-number differences between individuals were explained by variation in the length of the *apolipoprotein (a)* gene, which is highly polymorphic in the human genome because of variation in the number of copies of a 5.5-kb sequence encoding kringle repeats (Kamboh et al. 1991; Lackner et al. 1993). Indeed, this gene is located within the BAC affected in this patient (RP11-43B19), and therefore it is highly likely that this copy-number change also represents a genomic polymorphism.

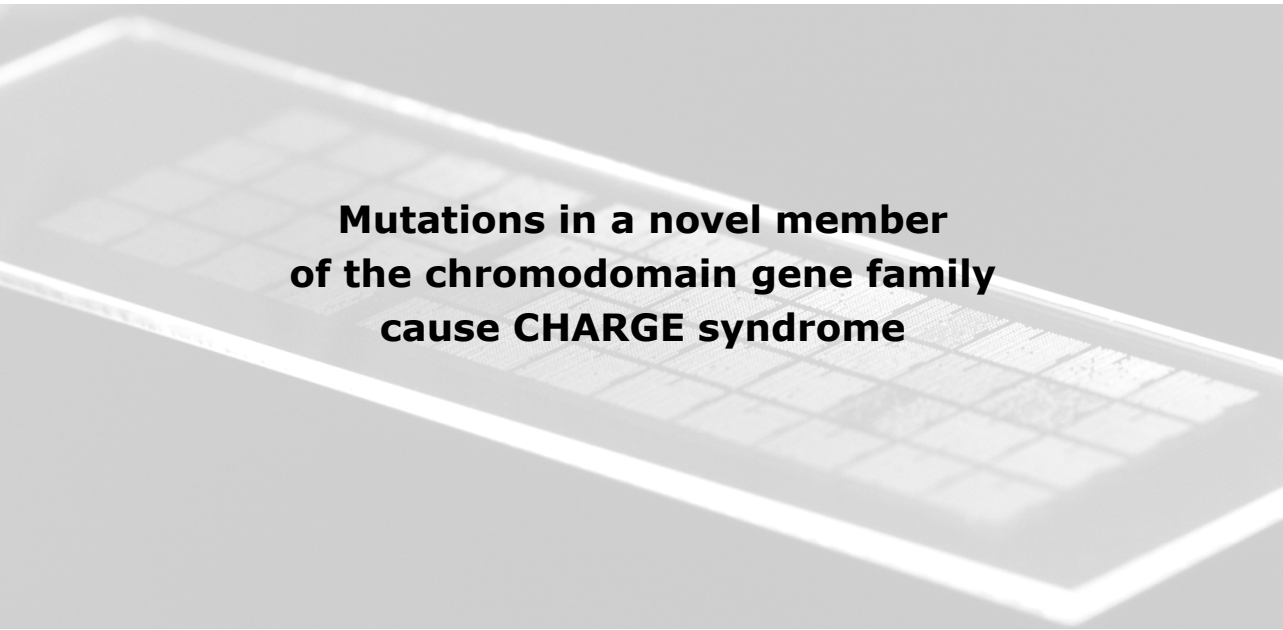
The detection of five microdeletions/-duplications in a series of 20 patients with mental retardation of unknown etiology in this pilot study underscores the strength of the array CGH technique. It should be mentioned that the patients in this study were selected on the basis of a phenotype suggestive of a chromosomal abnormality (de Vries et al. 2001). Therefore, this cohort may not be representative of the population of individuals with mental retardation as a whole. Also, two of the genomic abnormalities were identified in one of the parents as well, reducing the chance that these abnormalities are underlying the disorder. On the other hand, the resolution of the current microarray is, on average, 1 Mb, and, therefore, cases with more subtle anomalies may have been missed. To this end, we and other groups are in the process of constructing microarrays completely covering the genome with an average resolution of ~50 kb (for more information, see the BACPAC Resources Center's Human BAC Minimal Tiling Set Website).

There are three main applications for using array CGH in patients with mental retardation and malformations. First, it is to be expected that the number of interstitial microdeletions/-duplications will be comparable or may even exceed the ~5% submicroscopic, subtelomeric rearrangements currently reported among individuals with mental retardation (de Vries et al. 2003). For this reason, this 1-Mb–resolution array is currently being evaluated in a diagnostic setting in our department. Ultimately, this array-based copy number screening may partially replace karyotyping in this patient group. Second, these studies will facilitate the detection of genes involved in physical and mental development. Several such genes have already been identified as a result of systematic deletion mapping using microscopic chromosomal abnormalities (Romeo et al. 1994; Belloni et al. 1996; Robinson et al. 2003). The high resolution of the array CGH method allows for rapid and precise mapping of candidate genes for specific malformations. Third, a systematic analysis of genomic polymorphisms will give more insight into the flexibility and plasticity of the human genome. This latter may prove to be a fruitful field of study, given the fact that such large rearrangements will often involve multiple genes and may therefore serve as predisposing factors for multifactorial disorders.

Acknowledgements

We thank Yvonne Jonkers, Martin Elferink, and Monique Kersten-Niessen, for expert technical assistance; Ben Hamel, for referring cases; and Rolph Pfundt, for useful discussions. B.B.A.d.V. was supported by a grant from the Netherlands Organization for Health Research and Development (ZonMW).

Chapter 3



Mutations in a novel member of the chromodomain gene family cause CHARGE syndrome

Lisenka ELM Vissers, Conny MA van Ravenswaaij, Ronald Admiraal, Jane A Hurst, Bert BA de Vries, Irene M Janssen, Walter A van der Vliet, Erik HLPG Huys, Pieter J de Jong, Ben CJ Hamel, Eric FPM Schoenmakers, Han G Brunner, Joris A Veltman & Ad Geurts van Kessel

Nature Genetics 36: 955-957, 2004

Abstract

CHARGE syndrome is a common cause of congenital anomalies affecting several tissues in a nonrandom fashion. We report a 2.3 Mb *de novo* overlapping microdeletion on chromosome 8q12 identified by array comparative genomic hybridization in two individuals with CHARGE syndrome. Sequence analysis of genes located in this region, detected mutations in the gene *CHD7* in 10 of 17 individuals with CHARGE syndrome without microdeletions, accounting for the disease in most affected individuals.

Introduction

CHARGE syndrome denotes a nonrandom pattern of congenital anomalies including choanal atresia, malformations of the heart, inner ear and retina (Pagon et al. 1981). With an estimated birth incidence of 1:12,000 CHARGE syndrome is a common cause of congenital anomalies (Kallen et al. 1999). Most cases of CHARGE syndrome are sporadic, but several aspects of this condition, including the existence of rare familial cases and a high concordance rate in monozygotic twins, support the involvement of a genetic factor. Cytogenetic abnormalities have been described previously, but no specific locus has been identified (Lalani et al. 2003). Systematic genome scans by conventional comparative genomic hybridization (CGH) (Sanlaville et al. 2002) or microsatellite analysis (Lalani et al. 2003) did not identify a common genetic anomaly; neither did sequencing of candidate genes *PAX2* (Tellier et al. 2000) and *PITX2* (Martin et al. 2002).

Material and Methods

Individuals with CHARGE syndrome

Eighteen individuals with CHARGE syndrome and normal karyotypes upon routine cytogenetic investigation were selected based on the presence of at least 4 of the following 7 features: coloboma of the eye, heart malformation, choanal atresia, psychomotor and/or growth retardation, urogenital abnormalities, dysmorphic ears and/or deafness, and anomalies of the os petrosum (agenesis of the semicircular canals) on a CT-scan. Based on the initial array CGH results an additional, previously published individual with CHARGE syndrome and a *de novo* t(6;8) was included in this study (individual 2 in **Table 3.1**) (Hurst et al. 1991).

Array-based comparative genomic hybridization

Array CGH was performed essentially as described previously (Veltman et al. 2002; Vissers et al. 2003). Our genome-wide array contains 3569 BAC clones with an approximate average spacing of 1 clone per megabase (Vissers et al. 2003). The chromosome 8 array was composed of 918 BACs resulting in an average of 1 clone per 159 kb. These clones are part of a tiling resolution clone set (Ishkanian et al. 2004). Clones were spotted in triplicate onto UltraGAPS slides (Corning) using an Omnigrid 100 arrayer (Genomic Solutions). The array CGH profiles were established through co-hybridization of 500 ng Cy3-dUTP and Cy5-dUTP (Amersham Biosciences) labeled test and reference DNA, using a GeneTac Hybridization station (Genomic Solutions). After scanning, test-over-reference ratios (T/R values) were determined for each clone, $\log_2$-transformed and normalized per

array subgrid by Lowess curve fitting (Cleveland 1979; Vissers et al. 2003) Thresholds for copy number gain and loss were set at $\log_2$ T/R values of 0.3 and −0.3, respectively.

Mutation analysis

All coding sequences and flanking intronic sequences of the genes (*MGC39325, SDCBP, AL831990, CYP7A1, NSMAF, TOX, CA8, RAB2* and *CHD7*) located in the critical 8q12 interval were amplified and subsequently sequenced using oligonucleotides annealing 100-200 bp from the intron-exon boundaries. Primer sequences and PCR conditions are available upon request. All mutations were confirmed by bi-directional sequencing and checked for their *de novo* occurrence through mutation analysis of parental DNAs, if available.

FISH validations

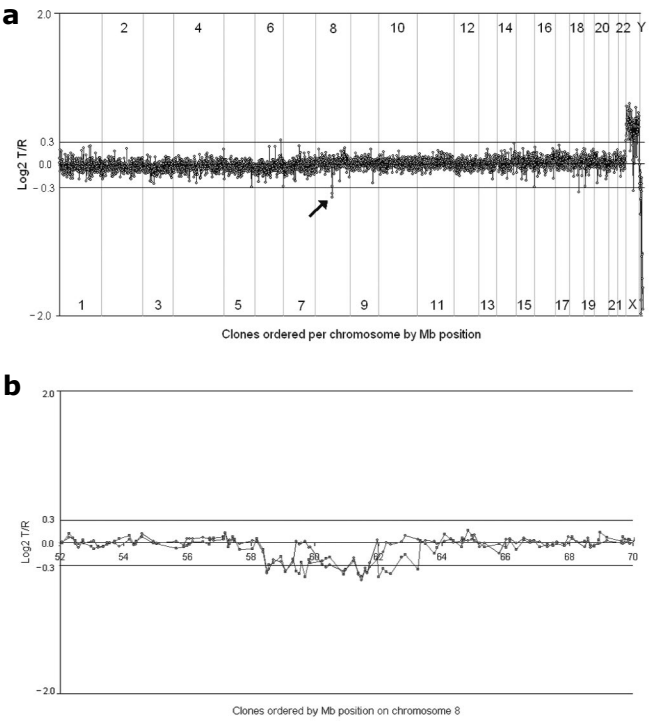
FISH validation experiments were performed on metaphase and interphase cells prepared from individual-derived and normal control-derived lymphoblastoid cell lines, respectively, using routine procedures. Probe labeling, slide preparation, hybridization and analysis were carried out according to routine procedures (Vissers et al. 2003).

Expression profiling by semiquantitative RT-PCR

For semiquantitative RT-PCR (Luijendijk et al. 2003), primers spanning intronic sequences were used. The expression profile of the *CHD7* gene was determined in 12 fetal tissues (eye, cochlea, brain, central nervous system, stomach, intestine, skeleton, heart, tongue, kidney, lung and liver) and 7 adult tissues (retina, cornea, brain, skeletal muscle, heart, kidney and lung). RT-PCR was performed in a volume of 35 µl. Samples of 8 µl were taken from the PCRs after 25, 30 and 35 cycles. Primer sequences and PCR conditions are available upon request.

Figure 3.1

Array CGH profiles of two individuals with CHARGE syndrome and FISH validation. (**a**) Array CGH genome-wide profile of the index individual with a copy number deletion of three adjacent clones on 8q12 (arrow). This profile represents the result of a single hybridization experiment; analysis of the replicate experiment identified only the 8q12 clones as being reproducibly deleted. Vertical lines indicate chromosome boundaries. (**b**) Profile of the same individual (brown squares) on the tiling chromosome 8 BAC array with 31 clones characterizing the deletion, and the chromosome 8 profile of an individual (black circles) with an apparently balanced t(6;8) translocation that overlaps with the deletion of the index individual. (**a**, **b**) Clones are ordered on the *x* axis according to physical mapping positions; $\log_2$ transformed test-over-reference ratios for each clone are given on the *y* axis. (**c**-**g**) FISH validation of the deletion in individual 1 (**c**) and testing for *de novo* occurrence in the accompanying parents (**d-e**) with clone RP11-91I20. (**f**) Determination of proximal deletion boundary (RP11-310L18 in red is normal; RP11-44D19 in green is deleted) and (**g**) determination of distal deletion boundary; RP11-274C23 in green is deleted whereas RP11-252M13 in red is normal. (**h-k**) FISH validation of the two distinct deletions in individual 2 and determination of the deletion boundaries. (**h**) The proximal deletion contains RP11-44D19 (green) on the centromeric site but not RP11-310L18 (red), and on the telomeric site (**i**) RP11-696A12 (red) but not RP11-630K16 (green). (**j**) Normal region between the two deletions is indicated by RP11-360N16 (red), whereas RP11-91I20 (green) marks the distal deletion. (**k**) This distal deletion contains RP11-265K14 (red) on the telomeric site, but not RP11-317H6 (green).


a
2.0
0.3
0.0
-0.3
-2.0
Log2 T/R
2
4
6
8
10
12
14
16
18
20
22
Y
1
3
5
7
9
11
13
15
17
19
21
X
Clones ordered per chromosome by Mb position
b
2.0
0.3
0.0
-0.3
-2.0
Log2 T/R
52
54
56
58
60
62
64
66
68
70
Clones ordered by Mb position on chromosome 8
c
d
e
f
g
h
i
j
k

Results and Discussion

We recently optimized array CGH (Pinkel et al. 1998) for high-resolution genome-wide screening of submicroscopic copy number changes (Vissers et al. 2003), and used this approach to identify microdeletions or duplications underlying CHARGE syndrome. We co-hybridized genomic DNA from two individuals with CHARGE syndrome with normal reference DNA onto a genome-wide BAC array with 1 Mb resolution. The genome-wide array CGH profile from one of these individuals is shown in **Figure 3.1a**. The only clones reproducibly deleted in this individual map to chromosomal band 8q12 and encompass a genomic interval of ~5 Mb. We confirmed the deletion by fluorescent *in situ* hybridization (FISH) analysis and proved that it occurred *de novo* (**Figure 3.1c-e**). The second individual with CHARGE syndrome included in this pilot study had no microdeletion or microduplication.

To further characterize the deletion in the index individual and to screen additional individuals for abnormalities of chromosome 8, we established a tiling resolution chromosome 8 array containing 918 overlapping BAC clones. After hybridizing DNA from the index individual onto this array (**Figure 3.1b**), we detected a deletion of 31 overlapping clones spanning a region of 4.8 Mb on 8q12, extending from RP11-44D19 to RP11-274C23 (**Figure 3.1f-g**). Notably, an individual with CHARGE syndrome with an apparently balanced chromosome 8 translocation was previously reported (Hurst et al. 1991). Hybridization of genomic DNA from this person onto the chromosome 8 BAC array detected two microdeletions overlapping with the one that we identified in the index individual (**Figure 3.1b**): one encompassing 6 overlapping clones (from RP11-44D19 to RP11-661A3, ~0.8 Mb) and one encompassing 11 overlapping clones (from RP11-51L11 to RP11-113D4, ~1.5 Mb). Between these two deleted regions, 6 clones (~0.9 Mb) showed normal test-over-reference ratios. We used metaphase FISH analysis to confirm the presence of and to determine the boundaries of the two distinct microdeletions in this person (**Figure 3.1h-k**). Although we could not verify *de novo* occurrence of the microdeletions in this case, the translocation was previously shown to be *de novo* (Hurst et al. 1991). Using data from these two individuals, we defined a shortest region of deletion overlap encompassing 2.3 Mb of genomic sequence on 8q12 (**Figure 3.2a**). We then screened 17 additional individuals with CHARGE syndrome using the chromosome 8 tiling array and detected no additional chromosome 8 copy number changes. Next, we sequenced the coding regions and the intron-exon boundaries of all nine annotated or predicted genes located in or just outside the shortest region of deletion overlap (**Figure 3.2a**). We identified ten heterozygous mutations in the gene *CHD7*, including seven stop-codon mutations, two missense mutations and one mutation at an intron-exon boundary (**Figure 3.2b-c, Table 3.1**). *CHD7* consists of 38 exons and has a genomic size of 188 kb. The stop-codon mutations were scattered throughout the gene: two in exon 2, one in exon 3, one in exon 26, two in exon 30 and one in exon 35. The two *de novo* missense mutations are predicted to lead to the amino acid substitutions I1028V (in exon 12), and L1257R (in exon 15). We observed one *de novo* mutation 7 bp upstream of exon 26 that could possibly affect splicing (IVS26-7G→A).

cDNA clone KIAA1416 (Nagase et al. 2000), encompassing the last 34 exons of *CHD7*, is a member of the chromodomain helicase DNA-binding (CHD) genes. CHD proteins belong to a superfamily of proteins that have a unique combination of functional domains, including

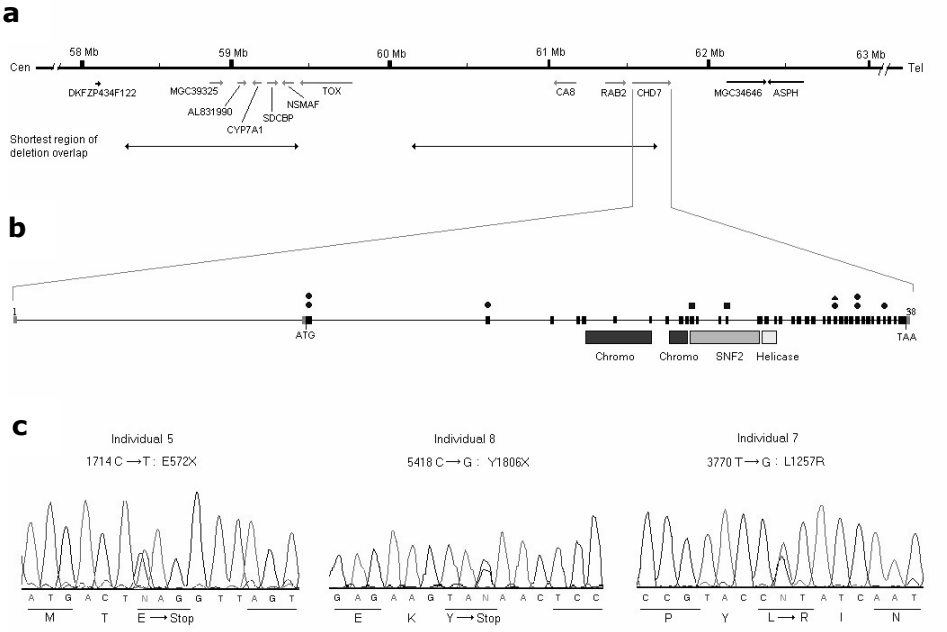


Figure 3.2
Detailed genomic view of 8q12, organization of the *CHD7* gene and mutations detected. (**a**) Transcript map of the deleted 8q12 genomic region. The shortest region of deletion overlap in the two individuals is shown. Genes in green were screened for mutations. Cen, centromeric; Tel, telomeric. (**b**) Genomic structure of *CHD7* indicating the positions of seven nonsense mutations (circles), two missense mutations (squares) and one intron-exon boundary mutation (triangle). The corresponding functional CHD7 domains are marked (colored bars). (**c**) Partial electropherograms obtained by direct sequencing of PCR products showing two nonsense mutations in individual 5 (1714C>T) and individual 8 (5418C>G) and one missense mutation in individual 7 (3770T>G).

two N-terminal chromodomains, a SNF2-like ATPase/helicase domain and a DNA-binding domain (Woodage et al. 1997) (**Figure 3.2b**). This class of proteins is thought to have pivotal roles in early embryonic development by affecting chromatin structure and gene expression (Cavalli et al. 1998). The congenital malformations found in CHARGE syndrome have their origin in early embryonic development. *CHD7* has ubiquitous expression in several fetal and adult tissues (**Figure 3.3**), including those affected in CHARGE syndrome.

The identification of seven heterozygous *CHD7* stop-codon mutations and two single copy 8q12 deletions of *CHD7* indicate that haploinsufficiency of this gene could account for most cases of CHARGE syndrome. Two amino acid changes are located in one of the functional domains of the CHD7, the SNF2 domain, and both affect a conserved amino acid (**Figure 3.4**). We did not observe any overt phenotypic difference between individuals with 8q12 deletions and those with nonsense or missense mutations in *CHD7* (**Table 3.1**). The index individual, with a 4.8 Mb deletion, has relatively severe mental retardation, which may be due to the presence of intronic or promoter mutations or to whole exon deletions. CHARGE

Table 3.1: Clinical characteristics of individuals with CHARGE syndrome and mutations in *CHD7*

		Major features						Associated features						
Individual	**Sex**	**Coloboma**	**Heart malformation**	**Atresia of Choanae**	**Retardation of growth and development**	**Genital hypoplasia**	**Ear abnormality including deafness**	**Vestibulo-Cochlear[a]**	**Facial nerve palsy**	**Cleft lip and/or palate**	**DNA mutation**	**Exon**	**Protein consequence**	**De novo**
1	F	+	+	-	+	-	+	+	-	-	del 8q12			yes
2[b]	F	+	+	-	+	-	+	U	+	-	del 8q12			U
3	F	+	+	-	+	-	+	U	-	+	469 C>T	2	R157X	U
4	F	-	+	+	+	-	+	U	-	-	1078 G>T	2	G360X	U[*]
5	M	-	-	-	+	+	+	+	+	+	1714 C>T	3	Q572X	yes
6	F	+	-	-	+	-	+	+	-	-	3082 A>G	12	I1028V	yes
7	F	+	-	-	+	+	+	+	-	-	3770 T>G	15	L1257R	yes
8	F	-	+	+	+	+	+	+	-	-	5418 C>G	26	Y1806X	yes
9	M	+	-	-	+	+	+	U	-	+	6051 T>A	30	C2017X	U
10	F	+	+	-	+	-	+	+	+	-	6070 C>T	30	R2024X	U[*]
11	F	-	+	+	+	-	+	+	-	-	7824 T>A	35	Y2608X	yes
12	M	+	+	-	+	+	+	+	-	+	IVS26-7G>A	26	Splice site	yes
13	F	+	+	+	+	+	+	+	-	+				
14	F	+	-	+	+	-	+	U	-	+				
15	M	+	-	-	+	-	+	+	-	-				
16	F	+	+	-	+	+	+	+	-	+				
17	F	-	+	-	+	+	+	+	+	-				
18	M	+	+	-	+	+	+	U	+	+				
19	M	+	+	+	+	+	+	+	-	-				

[a]: anomalies on CT-scan of inner ear (semicircular canal agenesis); [b]: Hurst et al. 1991; F: female. M: male. U: unknown. [*]: Only one parent available for testing, no mutation identified.

syndrome might also have a genetically heterogeneous etiology, as different genomic abnormalities have been identified in affected individuals (Sanlaville et al. 2002; Lalani et al. 2003). Notably, monosomy with respect to 22q11.2 has been reported to occur in individuals with features of both DiGeorge syndrome and CHARGE syndrome (de Lonlay-Debeney et al. 1997). Deletion of 22q11.2 was excluded in the individuals in this study who did not have *CHD7* deletions or mutations. Microdeletions encompassing the underlying gene have been reported to occur at low frequencies in single gene disorders (de Kok et al. 1995a; Johnson et al. 1998). We show that high-resolution genome-wide screening by array CGH is an effective new approach to localize such underlying genes. This approach is of particular interest for sporadic malformation syndromes that cannot be tackled by other mapping approaches because of reproductive lethality.

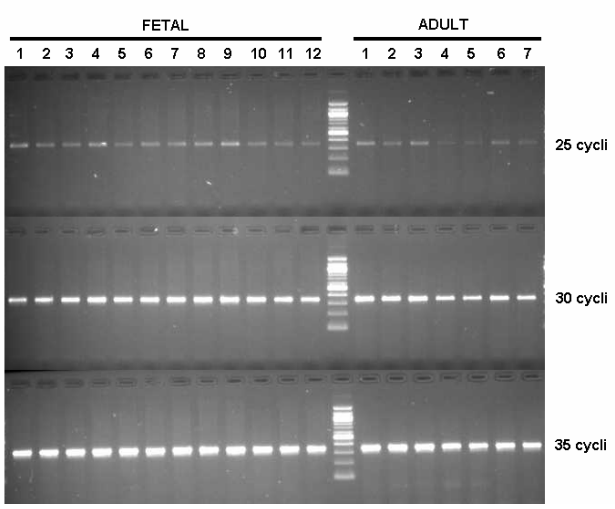


Figure 3.3
Expression profile analysis by semiquantitative RT-PCR analysis of *CHD7*. (**a**) RT-PRC specific for the transcript encoding the *CHD7* gene. Samples were taken after 25, 30 and 35 cycles. The RNA used for the RT-PCR was derived from the following tissues: fetal (1) eye, (2) cochlea, (3) brain, (4) central nervous system, (5) stomach, (6) intestine, (7) skeleton, (8) heart, (9) tongue, (10) kidney, (11) lung and (12) liver; adult (1) retina, (2) cornea, (3) brain, (4) skeletal muscle, (5) heart, (6) kidney and (7) lung. Expression was detected in all tissues at equal levels.

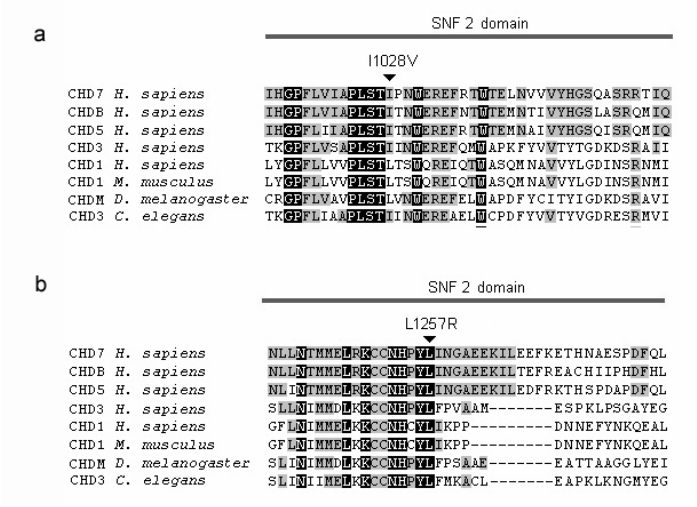


Figure 3.4
Conservation of mutated amino acids between CHD7 protein family in human and between different species. The affected residues are conserved within different proteins of the CHD protein family and between different species. (**a**) Mutation I1028V of individual 6 in exon 12 affecting the SNF2 domain. (**b**) Mutation L1257R of individual 7 in exon 15 also affecting the SNF2 domain. Black background: high conservation; Gray background: mild conservation; White background: no conservation.

Acknowledgements

We thank the affected individuals and their parents for their participation, H. Straatman and M. Leisink for statistical assistance, E. van Wijk and M. Sweers for expert technical assistance, E. Bongers for collecting samples from affected individuals, the sequence facility of our department and MACROGEN (Seoul, Korea) for sequencing, K. Osoegawa for participation in the production of the microarrays and M. Marra and M. Krzywinski for suggestions relating to the minimal BAC tiling set. This work was supported by grants from the Netherlands Organisation for Health Research and Development (B.B.A.d.V., H.G.B. and J.A.V.).

Accession numbers

Accession number of *CHD7* is XM_098762. The ORF encompasses 8994 bp and starts in exon 2 (http://www.ncbi.nlm.nih.gov/).

Chapter 4

A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism

David A Koolen, Lisenka ELM Vissers, Rolph Pfundt, Nicole de Leeuw, Samantha JL Knight, Regina Regan, R Frank Kooy, Edwin Reyniers, Corrado Romano, Marco Fichera, Albert Schinzel, Alessandra Baumer, Britt-Marie Anderlid, Jacqueline Schoumans, Nine V Knoers, Ad Geurts van Kessel, Erik A Sistermans, Joris A Veltman, Han G Brunner & Bert BA de Vries

Nature Genetics 38: 999-1001, 2006

Abstract

Submicroscopic genomic copy number changes have been identified only recently as an important cause of mental retardation. We describe the detection of three interstitial, overlapping 17q21.31 microdeletions in a cohort of 1,200 mentally retarded individuals associated with a clearly recognizable clinical phenotype of mental retardation, hypotonia and a characteristic face. The deletions encompass the *MAPT* and *CRHR1* genes and are associated with a common inversion polymorphism.

Introduction

Mental retardation is the most common developmental disorder, affecting intellectual and adaptive functions with a frequency of approximately 2–3% in the general population. Whole-genome scanning technologies such as array-based comparative genomic hybridization (array CGH) (Pinkel et al. 1998) have enabled the detection of interstitial submicroscopic copy number alterations in ~10% of individuals with mental retardation of unknown etiology (de Vries et al. 2005).

Material and Methods

Individuals

We tested 360 mentally retarded individuals for copy number changes using our genome-wide tiling path resolution microarray in a diagnostic setting (de Vries et al. 2005). Previous routine chromosome analysis was normal and subtelomeric MLPA failed to reveal any anomalies (SALSA MLPA kit P036, MRC Holland, Amsterdam, the Netherlands) in all individuals. In addition, we tested a cohort of 840 mentally retarded individuals for copy number changes in the 17q21.31 region by MLPA. DNA samples of these individuals were collected from centres in Oxford, UK (*n=130*), Antwerp, Belgium (*n=130*), Zurich, Switzerland (*n=100*), Troina, Italy (*n=130*), Stockholm, Sweden (*n=70*) and Nijmegen, The Netherlands (*n=280*). We obtained informed consent from the parents, including consent to publish photographs of all subjects described in this study.

Clinical description of individuals with the 17q21.31 deletion

Individual 1

The chromosomes of this 3 year-old-girl were analyzed prenatally and found to be normal. She was born at 35 2/7 weeks gestation with a birth weight of 2078 g (10th centile) and head circumference of 30.4 cm (5th centile). She was severely hypotonic and her development was considerable delayed: sitting at 3 years and no words at 3 years and 1 month. Magnetic Resonance Imaging (MRI) of the brain at 1 year of age showed widened ventricles and periventricular white matter changes. Electromyography (EMG) was normal, but Brainstem Evoked Response Audiometry (BERA) was abnormal. At 3 months she was treated for a congenital hip dysplasia. Diagnostic testing including DNA analysis for myotonic dystrophy, Spinal Muscular Atrophy (SMA), Prader-Willi syndrome, metabolic screening in blood and urine revealed no abnormalities.

On physical examination at the age of 3 years and 1 month, a height of 95 cm (40th centile), and a head circumference of 49.5 cm (50th centile) was noted. She had a long

hypotonic face with frontal bossing and bitemporal narrowing, ptosis, blepharophimosis, upward slanting palpebral fissures, epicanthal folds, large low set ears with hypoplastic crus superior, low nasal bridge, bulbous nasal tip, long columella, triangular nostrils, high palate with broad gums and a broad chin (**Figure 4.1a**). Her broad thorax had wide-space nipples and a mild pectus excavatum. She had long fingers, narrow long feet with long toes and hyperlaxity of the joints (Beighton score of 6/8). She was good-natured with normal eye contact.

Individual 2
This 17-year-old, moderately mentally retarded woman was born at 38 weeks gestation in breech position with a caesarian section and she had a birth weight of 1980 g (<3rd centile). In the first 3 weeks she received nasal catheter feeding because of low glucose levels and phototherapy because of hyperbilirubinaemia. She was hypotonic and started walking and speaking after the age of 2 years. She had an IQ of 48 points and attended special schooling. From the age of 1.5 to 3.5 years she had epileptic insults for which antiepileptic drugs were used. An MRI of the cerebrum showed wide ventricles, especially of the temporal horn of the lateral ventricles. In addition, routine chromosome analysis, metabolic screening in blood and urine and EMG, revealed no abnormalities. She developed a scoliosis at 13 years of age. She had normal hearing but mildly impaired vision (+4/+4).

On physical examination at the age of 17 years, low-normal height of 160.8 cm (10th centile), head circumference of 53 cm (10th centile) and weight 51.2 kg (50th centile for height) was noted. She had a long hypotonic face with ptosis, blepharophimosis, upward slanting palpebral fissures, large ears with hypoplastic crus superior, tubular pear-shaped nose with high nasal bridge and long columella, short philtrum, two missing upper teeth, everted lower lip and broad chin (**Figure 4.1b**). She had a thoracal scoliosis with a lumbar hyperlordosis. Her hands and fingers were long and slender with a simian crease in the left palm. Her feet had high arches with hallux valgus bilaterally and mild hammer toes. The lower part of the limbs was slender and there was a mild general decrease of strength. In addition, she had mild hyperlaxity of the finger joints (Beighton score of 2/9), and numerous moles on the skin. She had nasal speech and an amiable nature.

Individual 3
This 26-year-old, moderately mentally retarded male was born after an uneventful pregnancy at term with a normal birth weight of 3120 g (50th centile) but a large head circumference of 37.5 cm (>97th centile). He was notably hypotonic and computed tomography imaging of the brain at 4 months of age showed a communicating hydrocephaly without increased intracranial pressure. Both Somatosensory Sensory Evoked Potentials (SSEP) and BERA studies were abnormal at 6 months but normalized later in life. His development was retarded, walking at 3 years of age and he attended special school. He was operated on inguinal hernia and cryptorchidism. A scoliosis developed at 13 years of age. Diagnostic test including routine chromosome, FMR1 analysis and metabolic screening in blood and urine revealed no abnormalities.

On physical examination at the age of 26 years, low-normal height of 173 cm (10th centile), and a large head circumference of 63 cm (>97th centile) was noted. He had a long narrow face with a high, broad forehead, blepharophimosis, strabismus divergence, large ears with hypoplastic crus superior and large lobules, tubular pear-shaped nose with

Figure 4.1
Clinical characteristics of the 17q21.31 deletion syndrome. (**a**) Patient 1, at 3 years of age. (**b**) patient 2, at 17 years of age, and (**c**) patient 3, at 26 years of age. Note the characteristic facial features, a hypotonic face with ptosis, blepharophimosis, large low set ears, bulbous nasal tip, long collumella with hypoplasic alae nasi, and a broad chin.

bulbous tip and long columella, high palate with broad gums and diastemia frontal upper teeth, and a large broad chin (**Figure 4.1c**). The thorax was flat and broad with wide-spaced nipples and a scoliosis. His hands and fingers were long and hyperlax (Beighton score of 4/9). His feet had high arches with hallux valgus bilaterally and his skin revealed numerous moles. He had nasal speech and a friendly nature.

Array-based comparative genomic hybridization
Microarray preparation, hybridization, and data analysis were performed as described previously (de Vries et al. 2005). In brief, the array contains 32,447 BAC clones resulting in a complete coverage of the human genome (Ishkanian et al. 2004). The array CGH profiles were established through co-hybridization of 500 ng Cy3-dUTP labeled patient DNA and 500 ng sex-mismatched Cy5-dUTP labeled (Amersham Biosciences) reference DNA. After scanning, test-over-reference ratios were determined for each clone, $\log_2$-transformed and normalized by subtracting its local mean $\log_2$ test-over-reference ratio obtained by a weighted median filter. The normalized ratios were analyzed for loss and gain regions by a standard hidden Markov model (HMM).

Copy number screening of the 17q21.31 region by MLPA
We used MLPA (Schouten et al. 2002) to screen for copy number changes at 17q21.31 in the cohort of 840 mentally retarded individuals. A set of uniquely sized MLPA probes, hybridizing to exons 3 and 13 of *CRHR1*, and to exon 2 of *MAPT* was used. The probes were designed according to a protocol provided by MRC-Holland (http://www.mlpa.com/index.htm). The three MLPA probes were combined in one MLPA assay in conjunction with three standard control probes (*VIPR2*, *MRPL41*, and *KIAA0056*). Probe sequences are provided in **Table 4.1**. Hybridization, ligation and amplification of the MLPA probes were performed as described before (Schouten et al. 2002). Amplification products were identified and quantified by capillary electrophoresis on a genetic analyzer (ABI 3730 or 3100), using GeneMapper software (Applied Biosystems).

Table 4.1: MLPA primer sequences

Probe name	5'half-probe sequence	3'half-probe sequence
CRHR1 ex_13	CCAACCCGTGTCAGCTTTCACAGCA	TCAAGCAGTCCACAGCAGTCTGAGC
VIPR2 ex_12*	CGCGCCCAGTCCTTCCTGCAAACGGAG	ACCTCGGTCATCTAGCCCCACCCCTGC
MAPT ex_2	GGACAGGAAAGATCAGGGGGGCTACACCA	TGCACCAAGACCAAGAGGGTGACACGGAC
CRHR1 ex_3	CCCTGCCCTGCCTTTTTCTATGGTGTCCGCTACAATACC	ACAAGTAAGGAAGAAGTGGAGGGTGGACCATCTGCTGGG
MRPL41 ex_1*	GACCCTGACAACCTGGAAAAGTACGGCTTCGAGCC CACACAGGAG	GGAAAGCTCTTCCAGCTCTACCCCAGGAACTTCCTGCGCTAGCTG
KIAA0056 ex_8*	CAGCAATTATGCCAGCCTGACCTACCTTCAGATGGCTTGAAATGGTT	TACTACAGTCTGCATCACTATGTCTGAGACCCTTGTGTTCTCCATCC

* Control probes

Data were normalised by dividing each probe's signal strength by the average control probe signal strength of the sample. This normalised peak pattern was divided by the average peak height of all the samples in the same experiment. Copy number change detection was based on thresholds for gains and losses of 1.30 and 0.70 respectively (±3 SD). The MLPA analysis was repeated for all samples in which an aberration was identified. For these confirmation experiments, DNA samples of healthy controls were used for the normalization.

FISH validation studies and *de novo* occurrence

The 17q21.31 deletions identified by array CGH and MLPA were validated and tested for *de novo* occurrence by FISH analysis on fixed metaphase spreads from the respective patients and their parents as described before (de Vries et al. 2005). BAC clones in the aberrant 17q21.31 region were selected and used as probes in the FISH validation assay. In addition, parental DNA samples were tested using the same synthetic MLPA probe set described above.

Genotyping for H1 and H2, and parent-of-origin analysis

The presence of an intronic 238-bp deletion in intron 9 of the *MAPT* gene, characteristic for the H2 background (Baker et al. 1999), was used to genotype individuals with the 17q21.31 deletion and the accompanying parental DNA samples. The 238-bp deletion, if present, was determined by visualizing PCR product on an agarose gel. PCR reactions were performed using primer sequences GGAAGACGTTCTCACTGATCTG (sense) and AGGAGTCTGGCTTCAGTCTCTC (antisense) as described previously (Baker et al. 1999) in a 25 µl reaction mixture containing 50 ng of template DNA, 10 pmol of each primer, 0.4 mM dNTPs, 1.6 mM $MgCl_2$, and 1.25 U Amplitaq Gold Polymerase (Invitrogen). PCR conditions were as follows: 3 min 94°C followed by 5 cycles 94°C for 30s, 60°C for 30s, 72°C for 45 sec. Subsequently, a 10 cycle touchdown from 60°C to 50°C was performed, followed by 20 cycles 94°C for 30s, 50°C for 30s, 72°C for 45s, with a final extension 72°C for 10 min. The dinucleotide marker, *DG17S142* in intron 9 of the *MAPT* gene, was used for independent validation of the H1/H2 genotyping. Two additional markers, *D17S810* and *D17S920*, were used to study the parental origin of the deletions according to standard procedures.

Computational analysis

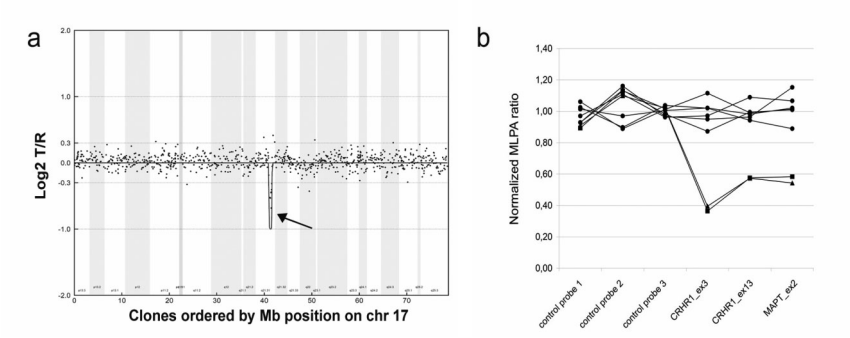
The chromosome 17 H1 lineage sequence map, was reconstructed for the region of interest (40.5 Mb – 42.0 Mb) based on the finished human chromosome 17 sequence. Interspersed repeat sequences within the reconstructed DNA sequence were eliminated by RepeatMasker (http://genome.ucsc.edu) and the repeat masked genomic sequence was analyzed using NCBI BLAST2 for the identification of LCRs (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html). All LCRs identified on H1 were in concordance with the previous study of Stefansson et al. (2005). LCR structures for the H2 lineage were adapted from Stefansson et al. (2005). The H2 lineage differs from the H1 lineage by a common 900-kb inversion polymorphism, the presence of a 32-kb DNA sequence (LCR17qE) that is present in two copies on the H2 lineage whereas only once on the H1 lineage (Stefansson et al. 2005), and the absence of LCR17qC.

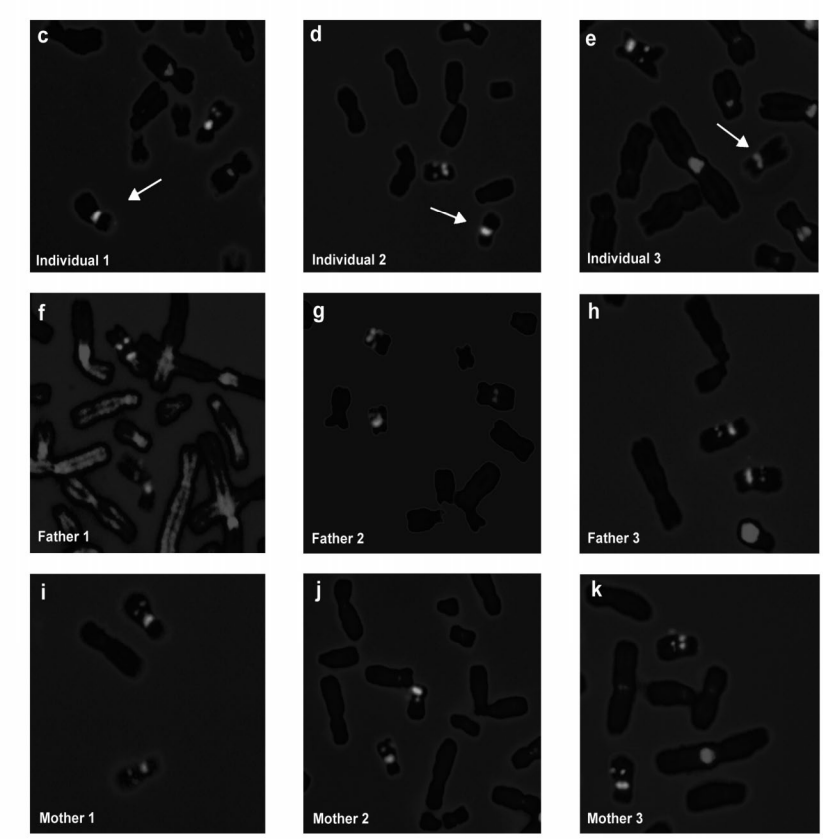
Results and Discussion

After obtaining informed consent, we tested 360 mentally retarded individuals within a diagnostic setting for copy number changes using our genome-wide tiling resolution microarray containing 32,477 BAC clones. In one individual, we identified a copy number loss with an approximate size of 600 kb at 17q21.31, encompassing eight BAC clones (**Figure 4.2a**). Subsequently, we screened a cohort of 840 mentally retarded individuals for deletions in the 17q21.31 region using multiplex ligation-dependent probe amplification (Schouten et al. 2002) with probes hybridizing to unique sequences in two genes located within this region, *MAPT* and *CRHR1*. By doing so, we identified two additional individuals with microdeletions in 17q21.31 (**Figure 4.2b**), with identical genomic sizes and deletion breakpoints at the BAC clone level. One of these breakpoints was identical to that of the index individual, whereas the other breakpoint was located ~100 kb distal to the first (**Figure 4.4**). For all individuals, we confirmed the presence of the 17q21.31 deletion by fluorescent *in situ* hybridization and showed that it arose *de novo* (**Figure 4.2c-k**). In addition, the 17q21.31 deletion region has not been reported to show copy number variation in normal individuals (Database of Genomic Variants; http://projects.tcag.ca/variation/).

We identified the deletions at 17q21.31 in a 3-year-old, moderately mentally retarded girl (individual 1), a 17-year-old, moderately mentally retarded woman (individual 2) and a 26-year-old, moderately mentally retarded male (individual 3). In all three individuals, severe hypotonia was present from birth onwards, leading to severely delayed motor development. None of the individuals could stand and/or walk before the age of three years. Upon physical examination, we noted characteristic facial features. All showed a long hypotonic face with ptosis, blepharophimosis, large, low-set ears, tubular pear-shaped nose with bulbous nasal tip, long columella with hypoplastic alae nasi and a broad chin (**Figure 4.1**). In addition, they all had long fingers, nasal speech and displayed an amiable and friendly disposition. The individuals had wide ventricles, as assessed by magnetic resonance imaging before the age of 1 year, whereas individual 1 had additional periventricular white matter abnormalities. Both individual 2 and individual 3 developed a hallux valgus, a pes cavus and, at 13 years of age, a thoracal scoliosis, probably due to marked hypotonia (**Table 4.2**).

Figure 4.2
Analysis of individuals with the 17q21.31 microdeletion syndrome. (**a**) Chromosome 17 array CGH profile of individual 1 with a copy number loss of eight adjacent BAC clones on 17q21.31 (arrow). Clones are ordered on the *x* axis according to physical mapping positions; $\log_2$-transformed test-over-reference (T/R) intensity ratios for each clone are given on the *y* axis. (**b**) Multiplex ligation-dependent probe amplification (MLPA) analysis, showing a deletion of two probes in *CRHR1* and one probe in *MAPT* for both individual 2 (triangles) and individual 3 (squares). The circles represent the MLPA ratios of healthy controls. FISH validation of the 17q21.31 deletion in individuals 1, 2 and 3 (**c-e**) and testing for *de novo* occurrence in the accompanying parents (**f-k**) using BAC clone RP11-656O14 (red) which is located within the deleted region. The centromere 17 probe was included for reference (green). All individuals show only one signal for RP11-6565O14 (arrow indicating aberrant chromosome 17). The *de novo* occurrence for the deletion was proven in all individuals.


a
Log2 T/R
Clones ordered by Mb position on chr 17
b
Normalized MLPA ratio
control probe 1
control probe 2
control probe 3
CRHR1_ex3
CRHR1_ex13
MAPT_ex2


c
Individual 1
d
Individual 2
e
Individual 3
f
Father 1
g
Father 2
h
Father 3
i
Mother 1
j
Mother 2
k
Mother 3

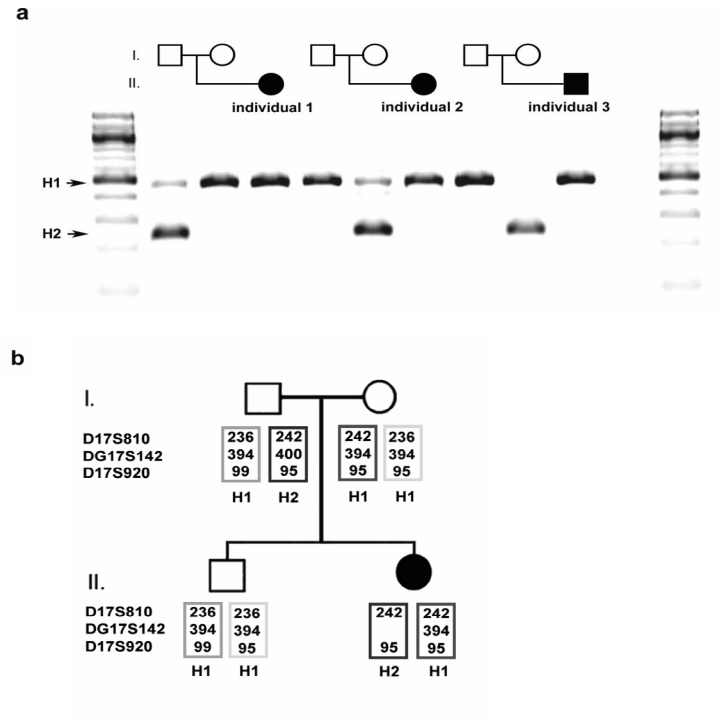


Figure 4.3

Genotyping for H1 and H2 lineage and parent-of-origin analysis. (**a**) Genotyping for the H1 (483 bp) and H2 (245 bp) lineages. Far left and far right: 100-bp marker lanes. Individual 3 carries the deletion on the H2 lineage and is of maternal origin. (**b**) Individual 1 and the parents were genotyped using *D17S810*, *DG17S142* and *D17S920* to determine the parental origin of the deletion. The deletion was of paternal origin, and was present on the H2 haplotype.

Notably, the deletions in 17q21.31 are all located within a genomic region that was recently reported to harbor a common 900-kb inversion polymorphism (Stefansson et al. 2005). For this region, two main and highly divergent haplotypes, designated H1 and H2, have been found. The H2 lineage, representing the 900-kb inversion polymorphism, is found at a frequency of 20% in Europeans and can be distinguished from the H1 lineage by genotyping of a dinucleotide marker (*DG17S142*) in intron nine of *MAPT* and by a characteristic 238-bp deletion in the same intron (Stefansson et al. 2005). For all three individuals, one of the parents carried the H2 haplotype: the father of individuals 1 and the mother of individual 2 were heterozygous for this haplotype, and the mother of individual 3 was homozygous (**Figure 4.3**). Parent-of-origin analysis showed that the deletion occurred on the H2 haplotype in individual 1 and 3, whereas the results for individual 2 were inconclusive (**Figure 4.3**). Notably, the H2 haplotype differs from the H1 by a directly oriented low-copy repeat (LCR), LCR17qE (**Figure 4.4**) that immediately flanks the breakpoints in all three individuals. This suggests that these deletions have resulted from nonallelic homologous recombination (Lupski 1998), mediated by this H2-specific LCR. Consequently, carriers of the H2 lineage are likely to be predisposed to nonallelic

Table 4.2: Clinical data from individuals with a 17q21.31 deletion

	Individual 1	Individual 2	Individual 3
Level of MR	Moderate	Moderate	Moderate
Hypotonia	+++	++	+++
Characteristic face			
Long	+	+	+
Blepharophimosis	+	+	+
Ptosis	+	+	+
Tubular pear-shaped nose	+	+	+
Long columella/ hypoplasic alae nasi	+	+	+
Broad nasal tip	+	+	+
Large ears	+	+	+
Broad chin	+	+	+
Friendly/amiable behavior	+	+	+
Brain anomalies	WV, PWM	WV	WV

MR: mental retardation; +: present; -: absent. Hypotonia: +: mild; ++: moderate; +++: severe. Brain anomalies: WV: wide ventricles; PWM: periventricular white matter.

homologous recombination, similarly to predisposing inversion polymorphisms, which have been observed in other microdeletion syndromes such as Williams-Beuren syndrome, Angelman syndrome and Sotos syndrome (Feuk et al. 2006).

We detected the three individuals with a similar 17q21.31 microdeletion in an unselected European cohort of 1,200 individuals with mental retardation, resulting in an overall detection frequency of 0.3%. As mental retardation occurs in 2–3% of the general population, it can be estimated that the prevalence of this new syndrome is between 1 in 13,000 and 1 in 20,000. This makes it less common than the estimated population prevalence of the 22q11 deletion (DiGeorge-velocardiofacial syndrome) of 1 in 4,000 (Lindsay 2001) but similar in frequency to Williams-Beuren syndrome, which has a frequency of 1 in 10,000 to 1 in 20,000 (Donnai & Karmiloff-Smith 2000). Notably, two other single cases have been reported (Shaw-Smith et al. 2004; Varela et al. 2006).

The 17q21.31 deletion encompasses two known genes, *CRHR1* (NM_004382) and *MAPT* (NM_005910), and at least two putative genes, *IMP5* (NM_175882) and *STH* (NM_001007532). Loss-of-function mutations have not been reported for any of these genes in humans. However, gain-of-function mutations in *MAPT*, encoding the microtubule-associated protein TAU, cause autosomal dominant forms of frontotemporal dementia and parkinsonism (Hutton et al. 1998; Rademakers et al. 2004). Abnormal filamentous TAU deposits have been reported as a pathological characteristic in several other neurodegenerative diseases (Rademakers et al. 2004). Haploinsufficiency for the microtubule-associated protein TAU may affect axonal elongation and neuronal migration, thereby explaining the major clinical features observed in the 17q21.31 microdeletion-positive individuals (that is, severe hypotonia and moderate mental retardation). In support of this, Tau-deficient mice showed muscle weakness and memory disturbance (Ikegami et al. 2000; Takei et al. 2000).

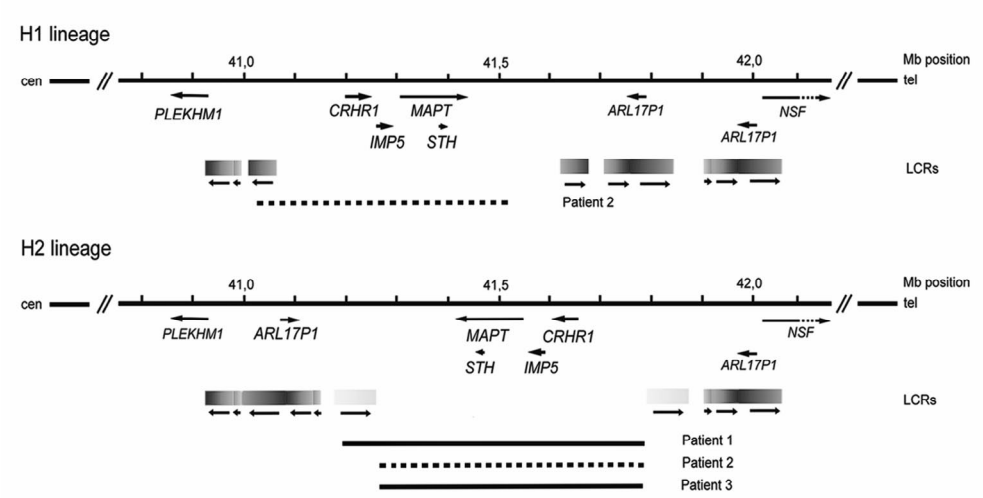


Figure 4.4
Transcript map and genomic architecture for the H1 and H2 lineage of the 17q21.31 region based on Cruts et al. (2005) and Takai et al. (2000). (Dotted) black lines represent the deleted 17q21.31 region identified in the three individuals. Genes located within this region are depicted by black arrows. The H1 and H2 lineage LCR17q structures are depicted as rectangles with colors signifying shared homology and horizontal arrows showing relative orientation (LCR17qA, red; LCR17qB, green; LCR17qC, purple; LCR17qD, blue; LCR17qE, yellow). Note the different genomic orientation of H2, the absence of LCR17qC and presence of LCR17qE. Cen, centromeric; Tel, telomeric.

In conclusion, we report a previously unknown, clinically recognizable syndrome defined by a 17q21.31 microdeletion that includes *MAPT*. The deletion is flanked by LCRs and is associated with a common inversion polymorphism. This observation underscores the relevance of genomic architectural features as the main determinant for the *de novo* occurrence of recurrent segmental aneuploidies.

Acknowledgements

We are thankful to all individuals and parents who participated in this study. In addition, we want to thank I. Scheltinga, I. Neefs, I. Janssen, W. van der Vliet, H. Mieloo, H. Venselaar, G. Schobers and W. Nillesen for expert technical assistance and C. van Ravenswaaij-Arts, M. Ruiter, J. de Bruin and K. Ultee for referral of individuals and clinical analysis. This work was supported by grants from the Netherlands Organisation for Health Research and Development (ZonMW) (D.A.K., J.A.V., H.G.B. and B.B.A.d.V.), Hersenstichting Nederland (B.B.A.d.V), The Health Foundation (S.J.L.K.) and The Swedish Research Council (J.S. and B.M.A.).

Chapter 5

Variation of CNV distribution in five different ethnic populations

Stefan J White*, Lisenka ELM Vissers*, Ad Geurts van Kessel, Renee X de Menezes, Ersan Kalay, Anna-Elina Lehesjoki, Pier C Giordano, Ellen van de Vosse, Martijn H Breuning, Han G Brunner, Johan den Dunnen & Joris A Veltman

In press, Cytogenetic and Genome Research, 2007

Abstract

Recent studies have revealed a new type of variation in the human genome encompassing relatively large genomic segments (~100 kb – 2.5 Mb), commonly referred to as copy number variation (CNV). The full nature and extent of CNV and its frequency in different ethnic populations is still largely unknown. In this study we characterized a set of twelve CNVs previously detected by array CGH. More than 300 individuals from five different ethnic populations, including three distinct European, one Asian and one African population, were tested for the occurrence of CNV using multiplex ligation-dependent probe amplification (MLPA). Seven of these loci indeed showed CNV, i.e., showed copy numbers that deviated from the population median. More precise estimations of the actual genomic copy numbers for (part of) the *NSF* gene locus, revealed copy numbers ranging from two to at least seven. Additionally, significant inter-population differences in the distribution of these copy numbers were observed. These data suggest that insight into absolute DNA copy numbers for loci exhibiting CNV is required to determine their potential contribution to normal phenotypic variation and, in addition, disease susceptibility.

Introduction

The genomic landscape shows marked variation in the distribution of a number of features such as single nucleotide polymorphisms (SNPs), various repeats such as LINEs and SINEs, microsatellite repeat polymorphisms and insertion/deletion polymorphisms. Recently, the use of genome-wide molecular methods such as array-based comparative genomic hybridization (array CGH) (Iafrate et al. 2004; de Vries et al. 2005; Sharp et al. 2005), representational oligonucleotide microarray analysis (ROMA) (Sebat et al. 2004), fosmid end-sequencing (Tuzun et al. 2005) and SNP arrays (Wirtenberger et al. 2006) have revealed the existence of an additional class of large-scale genomic variations. These variations may encompass genomic segments ranging in size from ~100 kb to ~2.5 Mb, and are commonly referred to as copy number variations (CNVs). In addition, ~1,000 intermediate length common deletion variants (<10 kb) have been identified and characterized in the human genome by taking advantage of technologies designed to catalogue SNPs present in the human population (Conrad et al. 2006, Hinds et al. 2006; McCarroll et al. 2006). Remarkably, comparison of DNA copy number variations identified in the studies mentioned above showed very little overlap (Eichler 2006). Even when regions were reported to show CNV in multiple studies, the frequencies at which they were reported often differed markedly.

The discrepancies observed between these studies may be due to differences in the technologies used to identify copy number variation, i.e., the resolution of the platforms used and the difficulty in discriminating between true variation and experimental noise due to lack of proper validation. In addition, the discrepancies may be related to the selection of normal control individuals used for assessing the CNVs. For insertion/deletion polymorphisms, for example, it is known that allele frequencies differ between populations with different ethnic backgrounds (Weber et al. 2002). As yet, there is almost no information available on the actual copy numbers at which different CNV regions are present within a given population. Because of the comparative nature of most array-based studies, regions are reported to show a relative loss or gain as compared to a reference

sample or a pool of references. Based on the variation present, however, the selection of references may affect the relative frequencies of gains or losses observed.

Although the mechanisms by which CNVs arise are largely unknown, it has been suggested that certain genomic features such as low copy repeats (LCRs) may be involved in the formation of common CNVs (Hind et al. 2006; McCarroll et al. 2006). LCRs are defined as genomic segments >1kb in size that occur in two or more copies per haploid genome, with the different copies sharing >90% sequence identity. These LCRs may mediate nonallelic homologous recombination (NAHR), a mechanism known to lead to the formation of deletions and duplications associated with genomic disorders (Inoue & Lupski 2002; Shaw & Lupski 2004). Additionally, gaps in the genome assembly, representing sequences with high repeat content including LCRs, may point at regions prone to CNV (Bailey et al. 2001; Zody et al. 2006).

For this study we selected twelve recently detected CNV regions, based on their reported high frequencies in the general population (Iafrate et al. 2004; de Vries et al. 2005; http://projects.tcag.ca/variation). We characterized these regions for genomic architectural features and used multiplex ligation-dependent probe amplification (MLPA; Schouten et al. 2002; White et al. 2004) to determine CNV frequencies in five ethnic populations, including three distinct European, one Asian and one African population, in total comprising more than 300 individuals. In addition, using these semi-quantitative MLPA results we attempted to estimate the actual DNA copy numbers present in the different populations tested. Our analyses underscore the presence of both intra- and inter-population differences in DNA copy number variation. These differences may be relevant for normal phenotypic variation and/or disease susceptibility.

Material and Methods

Individuals of different ethnic populations

DNA samples from normal, healthy individuals from five different ethnicities (Dutch (*n=84*), Turkish (*n=64*), Finnish (*n=36*), Indonesian (*n=88*) and African Pygmy (*n=37*)) were used to assess CNV distribution by multiplex ligation-dependent probe amplification (MLPA; see below). All DNA samples were isolated from peripheral blood leucocytes and were obtained through appropriate informed consent. The Dutch, Finnish and Turkish DNAs were isolated using a QIAamp kit (QIAgen, Valencia CA) according to instructions of the manufacturer. The remaining samples were isolated using standard desalting DNA isolation procedures.

Selection of regions with CNV

From the database of genomic variation (http://projects.tcag.ca/variation) twelve CNV regions were selected for assessment (**Table 5.2**). This selection was based on their estimated high frequencies in the general population (>5%). All CNVs were previously identified by array CGH (**Table 5.2**).

MLPA probe design

MLPA probe design was performed as described by Schouten et al. (2002), and White et al. (2004). The oligonucleotides were purchased from Illumina Inc. (San Diego, CA) or Isogen (IJsselstein, the Netherlands). At least two probes were selected within each

Table 5.1*:* MLPA Probe sequences

locus	name	5'half-probe sequence	3'half-probe sequence
1p32.3	DAB1a	CATTGTGTTTGAGGCTGGACACGA	GCCAATCCGTGATCCCGAAACGGAAGAA
1p32.3	DAB1b	GTCCGGTACAAAGCCAAATTGATCGGGAT	TGATGAAGTTTCCGCAGCTCGGGGAGACA
2q31.1	MYO3Ba	GCATTCACAAGCCCAGAGTTCTCCAAAAGGGTGCGA	TATCTTCGCAGGACATGCAAACAAGGTAGCTGGATA
2q31.1	MYO3Bb	GTAATTTGGCCCAGACAAGAGCTAGGATAACAGTGGCCTCAAGT	TCTTTGCCTCCACATTTCAGTGCTGGGAAAGCCAAG
2q35	DESa	GACGTGGATGCAGCTACTCTAGCT	CGCATTGACCTGGAGCGCAGAATTGA
2q35	DESb	CCTGAGCAAAGGGGTTCTGAGGTCCA	TACCAAGAAGACGGTGATGATCAAGACC
5q13.2	SMN1	GTAGTCCTGAAGCTTCCTAAGAAACTTCACATCAGGTGAT	GGATAGGAGCAACCCTGTAAAACCAGCCTTAGAC
5q13.2	SMN2	GCGATAACCACTCGTAGAAAGCGT	GAGAAGTTACTACAAGCGGTCCTCCC
7p21.1	TWIST1a	CGAGCTGGACTCCAAGATGGCA	AGCTGCAGCTATGTGGCTCACGAG
7p21.1	TWIST1b	CAGTGATTCCCAGACGGGCAGCGGCAGGATCCTCACACCTCT	GCATTCTGATAGAAGTCTGAACAGTTGTTTGTGT
7q35	ORF22a	GTCAGTGGGGCCTGGCATAAACTATTAGAG	AAATTCTCTGGGTTAACATCCAAGCTGGGAAC
7q35	ORF22b	GTCAGGCTGGCTTGTGTGGACACCTCCTCCAATGAGGC	TGCCATCATGGTGTCTAGCATTGTTCTTCTGATGAC
8p23.1	DEFB103B	GGAACCTTTATAAAGTGACCAAGCACACCTTTTCATCCAG	TCTCAGCGTGGGGTGAAGCCTAGCAGCTATGA
8p23.1	DEFB104A	GCAGCCCCAGCATTATGCAGAGACTT	GTGCTGCTATTAGCCATTTCTCTTCTACTC
8p23.1	DEFB4	GTGAGATGGGCCAGGGAAATAGGAGGGTTGGCCAAAT	GGAAGAATGGCGTAGAAGTTCTCTGTCTCCTCT
10q11.22	PPYR1	CTGAACATTGCCAGGATTCCGTGGACGTGATGGTCT	TCATCGTCACTTCCTACAGCATTGAGACTGTCGTGG
10q11.22	ANXA8	CGGCAGGCGAGAATATTCGTGGGACT	GATGAGATGAAATTCATCACCATCCTGTGCAC
10q11.22	SYT15	GGAGTTTGGCGACCTCCAGT	TCTGCCTCAGCTACAACGACTACCTG
14q11	ACTBL1	GGGCACTTCTGGAGACCACGACGATTCT	GCTATGAAGACACTCAGGAGCAAGAT
14q11	OR4Q3	CTTGTCTGCTTCTTGGTCTTACTATTCTCT	TATGCTATCATCCTGATCACCCTGAGAACA
17q12	TBC1D3a	CGGGAGATCAGCCGAAAGAGCAAGT	GGGTGGATATGCTGGGAGACTGGGAGAAATACAAA
17q12	TBC1D3b	GCAACCGGTTCGTTGATACCTGG	GCCAGGGATGAGGACACTGTGCTCAAG
17q21.31	NSF_12	CAAGAATGAGAGGGCATCAGTTACTCTCTGCTGATGTAGACATT	AAAGAACTGGCCGTGGAGACCAAGAATTTCAGTG
17q21.31	NSF_14	GGAGCTGCTGGTGCAGCAGACTAAGAACAGTGAC	CGCACACCATTGGTCAGCGTGCTTCTGG
17q21.31	NSF_2	GGCGGCAAGATGTCCTACAGATGAATTATCTTTAACCAATTGTGCAGT	TGTGAATGAAAAGGATTTCCAGTCTGGCCA
17q21.31	NSF_9	GAATGCAAGAGAGCCCAAAGTGGTCAATGGGCCAGAAATCCTTAACA	AATATGTGGGAGAATCAGAGGCTAACATTCGCA
22q11.21	PRODH	GAGATGGAGGAGCTGGGCCTGCATCCTGCTGACCACCGGGTGTACTTT	GGACAGCTGCTAGGCATGTGTGACCAGATCAG
22q11.21	GGT2	GGCTGGCCTTTGCCAGCATGTTCA	ACAGCTCGGAGCAGTCCCAGAAGG

genomic region, and exon sequences were preferentially chosen. Control probes were selected from known autosomal dominant disease genes (*CREBBP, EP300, EXT1, EXT2*), which would not be deleted or duplicated without an obvious phenotype. For genes classified as single copy in the genome browser, probes were designed such that they only recognized a single genomic sequence. For genes known to be present in duplicons, probes were designed such that they only recognized sequences within the selected region. Where significant homology to multiple regions was indicated by BLAT analysis (Kent 2002) an attempt was made to position the ligation site of the two oligonucleotides over a mismatch, thus allowing accurate discrimination from related sequences. The probe sequences are listed in **Table 5.1**.

MLPA assay and data analysis

The MLPA assay was carried out as described before (White et al. 2004). In short, for each MLPA reaction 100-200 ng genomic DNA was used, and the PCR reactions were performed for 31-33 cycles. Subsequently, the PCR products were analyzed on either an ABI 3700 or an ABI 3730 capillary sequencer (Applied Biosystems) and the obtained data were normalized as described before (White et al. 2004). Briefly, probes in a given colour were normalized against at least two control probes of the same colour. For each locus there was at least one probe in each colour, so each locus was in fact analyzed by two independent calculations.

The height of each probe-specific peak was divided by the sum of at least two control peaks to give a ratio. The median ratio for each locus across all samples per population within an assay was calculated and used to normalize the ratios around a value of 1. Initial analysis set thresholds for gains or losses of 1.25 and 0.75, respectively. A gain or loss was only scored if at least two probes within a given locus were outside the thresholds.

For validation of CNV detection by MLPA, normalized MLPA results for several individual CNV regions were $\log_2$ transformed and subsequently compared to array CGH results that were previously obtained by de Vries et al. (2005). For these validation studies, the same series of 100 patient DNA samples was used as described in de Vries et al. (2005). For the *DEFB*, *NSF* and *TBC1D3* loci, in which most variation was seen, further analysis was performed by averaging the value of all probes within each locus per sample.

Genotyping H1 and H2 on 17q21.31

The presence of a 238-bp deletion in intron 9 of the *MAPT* gene, characteristic for the H2 line, was used to genotype the Dutch individuals. The 238-bp deletion was determined through visualizing PCR products on agarose gels. PCR reactions were performed using primer sequences GGAAGACGTTCTCACTGATCTG (sense) and AGGAGTCTGGC-TTCAGTCTCTC (antisense) as described previously (Baker et al. 1999) in a 25 µl reaction mixture containing 50 ng of template DNA, 10 pmol of each primer, 0.4 mM dNTPs, 1.6 mM $MgCl_2$, and 1.25 U Amplitaq Gold Polymerase (Invitrogen). PCR conditions were as follows: 3 min 94°C followed by 5 cycles 94°C for 30s, 60°C for 30s, 72°C for 45s. Subsequently, a 10 cycle touchdown from 60°C to 50°C was performed, followed by 20 cycles 94°C for 30s, 50°C for 30s, 72°C for 45s, with a final extension 72°C for 10 min.

LCR identification by computational analyses

The presence and location of LCR structures were reconstructed using computational analysis. In brief, genomic sequences were downloaded from the UCSC genome browser and interspersed high copy repeat sequences were eliminated by RepeatMasker

(http://genome.ucsc.edu). Next, the repeat-masked genomic sequence was analyzed using NCBI BLAST2 for the identification of LCRs (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html).

Results

Selection, characterization and validation of recurrent CNVs

From the Database of Genomic Variants, we selected twelve genomic regions for the assessment of CNV in different ethnic populations. These regions, all recently detected by array CGH, were reported to show frequent CNV (http://projects.tcag.ca/variation) and varied in size from 119 kb to 2.57 Mb. Although for most regions multiple studies have indicated the presence of CNV, the frequencies at which CNV was detected differed markedly (**Table 5.2**). Since genomic architectural features such as LCRs and gaps in the genome assembly may be related to the presence of CNV at specific genomic locations, we evaluated the presence of such features for the twelve regions selected (**Table 5.2** and **Figure 5.1**). Gaps present within 100 kb of a CNV were documented for three of the selected regions, i.e., 8p23.1, 10q11.22 and 17q12. LCRs within or flanking a CNV were found to be present in eight of the selected regions, and insertion/deletion polymorphisms were reported in one of the selected regions (2q35). In three regions, combinations of gaps and LCRs were observed, thus contributing to the complexity of the regions under evaluation. In addition to LCRs and gaps, another genomic architectural feature was observed, namely a 900 kb inversion polymorphism within the 17q21.31 region (Stefansson et al. 2005). These data support the notion that these specific genomic architectural features may be related to the occurrence of CNV.

The presence of gaps and complex repeat structures within and/or in close proximity to loci exhibiting CNV could hamper MLPA probe design, thereby affecting the detection of CNVs in the individuals to be tested. Therefore, we first validated the MLPA performance for this type of analysis. In addition, the MLPA results for six of the twelve regions were compared to previously obtained array CGH results in a cohort of 100 patients with mental retardation (de Vries et al. 2005). **Figure 5.2** shows an example of such a correlation for the CNV locus at 17q12. The MLPA probes showed a consistent performance over two experiments. Furthermore, all patient samples showing gains and losses by array CGH also showed higher and lower MLPA values, respectively, for the probe(s) located within this CNV region. Similar results were obtained for the five other regions tested. From these data we conclude that the MLPA technique is suitable for the reliable detection of DNA copy number variation.

Assessment of selected recurrent CNVs in different ethnic populations

In order to assess the occurrence of recurrent CNVs in different ethnic populations, the twelve selected CNV regions were tested by MLPA in DNAs derived from 309 individuals from Dutch, Finnish, Turkish, Pygmy and Indonesian origin. All data were independently normalized per population, allowing the detection of inter- and intra-population differences. It is important to note that an MLPA value of 1.0 in this study represents the population-specific copy number median, and does not necessarily represent the usual standard of two copies. As a first analysis, the number of individuals with either a gain or a loss at each locus was determined. For this analysis, thresholds for relative gain and loss

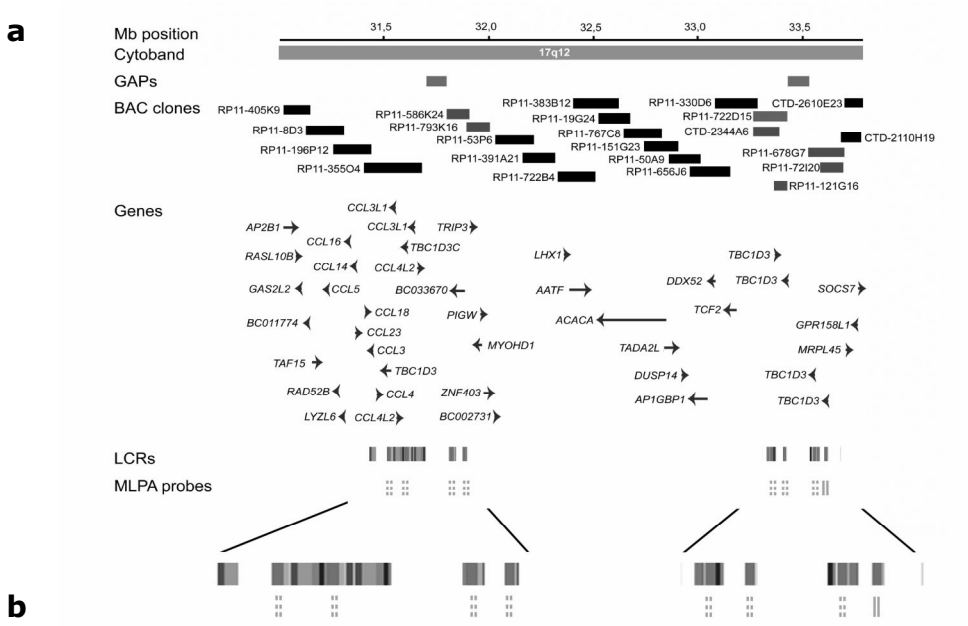
from the median of the population were set at 1.25 and 0.75, respectively. These values have been routinely used in MLPA screening for deletions and duplications in single copy loci. Using this approach, CNV was observed in all populations tested for seven out of twelve selected loci. CNV frequencies at these seven loci ranged from 3% to 52% (of all 309 individuals; **Table 5.3**). In the remaining five regions, however, no significant CNV was observed in any of the populations tested (**Table 5.3**).

For four of the seven regions showing moderate CNV, including 5q13.2, 10q11.22, 14q11 and 22q11.21, no significant difference was seen between populations for any of the loci tested (Bionominal test with normal approximation $p>0.05$). The regions exhibiting the highest degree of CNV within each population encompassed the beta defensin cluster (8p23.1; 26% of all individuals showed CNV), the *TBC1D3* gene (17q12; 24% of all individuals showed CNV), and the proximal part of the *NSF* gene (17q21.31; 52% of all individuals showed CNV).

In an attempt to estimate the actual copy numbers present for the three most variable loci within each population, we analyzed the MLPA results semi-quantitatively. This was achieved by calculating the average probe values within each locus per individual, which were then ranked from low to high. In addition, the percentiles 10%, 25%, 50%, 75% and 90% were calculated in order to observe inter-population differences in the distribution of the actual copy numbers (**Figure 5.3**). By doing so, we observed a ~3.5 fold change between the lowest and highest MLPA values for the beta defensin gene cluster. With a previously determined minimal copy number of two (Hollox et al. 2003), our data indicate a copy number range from two to seven in the majority of all populations. The 10% to 90% percentiles for the beta defensin gene region, however, differed by <10% over all populations, suggesting that the frequencies of the actual copy numbers are largely the same within each population. There were, however, clear differences in the percentiles between populations for *TBC1D3* (**Figure 5.3**). Significant differences can be found when performing Fisher's exact test on pairwise comparisons of population-specific discretized ratios using various cut-offs for the *TBC1D3* locus (1.3, 1.4, 1.5; $p<0.05$), while the same comparisons yielded no significant differences with the same discretizations for the beta defensin gene cluster. However, since we are not able to estimate absolute copy numbers, we cannot take this analysis for either the beta defesin gene cluster or the *TBC1D3* gene locus any further.

In contrast, subgroups of individuals displaying different *NSF* gene copy numbers were evident in all populations tested (**Figure 5.4a-b**). Subgroups were identified and

Figure 5.1
Examples of genome architecture and MLPA probe locations. (**a**), 17q12 (**b**) and 17q21.31. Note that for the 17q21.31, a 900-kb inversion polymorphism including the *NSF* locus has been observed in the general population, represented by the H1 and H2 orientation (Stefansson et al. 2005). For all regions, Mb position, cytoband, BAC clones, genes located within the region, LCRs/ GAPs (if present) and the location of the MLPA probes are depicted. The BAC clones that were previously reported to show CNV are colored red (Genome Variation Database). The LCR structures, if present, are depicted in colors to better represent their positional orientation with respect to each other. The locations of the MLPA probes are indicated by vertical orange solid lines, whereas dotted orange lines indicate alternative MLPA locations, sharing high homology to the region the probe was originally designed for.


a
Mb position
Cytoband
31,5
32,0
32,5
33,0
33,5
17q12
GAPs
BAC clones
RP11-405K9
RP11-8D3
RP11-196P12
RP11-355O4
RP11-586K24
RP11-793K16
RP11-53P6
RP11-391A21
RP11-722B4
RP11-383B12
RP11-19G24
RP11-767C8
RP11-151G23
RP11-50A9
RP11-656J6
RP11-330D6
RP11-722D15
CTD-2344A6
RP11-678G7
RP11-72I20
RP11-121G16
CTD-2610E23
CTD-2110H19
Genes
CCL3L1
AP2B1
CCL3L1
TRIP3
CCL16
TBC1D3C
RASL10B
CCL14
CCL4L2
GAS2L2
CCL5
BC033670
LHX1
AATF
DDX52
TBC1D3
TBC1D3
SOCS7
BC011774
CCL18
PIGW
ACACA
TCF2
GPR158L1
CCL23
CCL3
MYOHD1
TADA2L
MRPL45
TAF15
TBC1D3
DUSP14
TBC1D3
RAD52B
CCL4
ZNF403
AP1GBP1
TBC1D3
LYZL6
CCL4L2
BC002731
LCRs
MLPA probes
b
H1 Orientation
Mb position
Cytoband
41,0
41,5
42,0
42,5
17q21.31
BAC clones
RP11-656O14
RP11-798G7
RP11-733O21
RP11-397D10
RP11-782E1
RP11-86C1
RP11-162O14
RP11-193J3
RP11-57A24
RP11-141H9
RP11-79O18
RP11-63A1
RP11-67G6
RP11-502C16
Genes
ARHGAP27
CRHR1
PLEKHM1
STH
MAPT
LOC284058
ψ NSF
ARL17P1
NSF
WNT3
ARL17P1
WNT9B
CDC27
GOSR2
LCRs
MLPA probes
H2 Orientation
Mb position
Cytoband
41,0
41,5
42,0
42,5
17q21.31
Genes
ARHGAP27
ψ NSF
PLEKHM1
ARL17P1
LOC284058
MAPT
STH
CRHR1
NSF
WNT3
ARL17P1
WNT9B
CDC27
GOSR2
LCRs
MLPA probes

confirmed by applying hierarchical clustering using complete linkage. For estimations of the *NSF* copy numbers present in the populations tested, we extrapolated the findings of a previous study using an Icelandic population (Stefansson et al. 2005), which indicated that the minimum copy number present for this region is two. Based on the best-fit correlation coefficient all populations, apart from the Pygmy population, appeared to have copy numbers ranging from two to at least seven (**Figure 5.4c**). In the Dutch and Turkish populations the median copy number was three, whereas in the Finnish, Pygmy and Indonesian populations this median was four (**Figure 5.5**). The actual degree of CNV in four of the populations, i.e., individuals with a copy number deviating from the median, was virtually identical, ranging from 66% to 69%. In the Pygmy population only 49% of the individuals showed CNV. The relative frequency of different copy numbers between the five populations did show marked differences (Chi-square test; $p<0.001$) (**Figure 5.5**). A highly significant difference was seen between the Dutch and Indonesian populations (Chi-square test; $p<0.001$). The Indonesian population showed a skew towards higher copy numbers, with 40% of the individuals having five or more copies. This percentage was markedly different from the Dutch population, in which only 12% of individuals exhibited five or more copies. In contrast, there was no significant difference observed between the Dutch and Turkish populations (Chi-square test; $p=0.469$).

For the 900-kb inversion polymorphism at 17q21.31, containing the *NSF* gene, two main and highly divergent haplotypes, designated H1 and H2, have been unveiled (Stefansson et al. 2005). H2 can be distinguished from H1 by a characteristic 238-bp deletion in intron nine of the *MAPT* gene (Baker et al. 1999). In order to establish a correlation between the *NSF* copy numbers and the respective genotypes, individuals from the Dutch population were genotyped for these haplotypes. Previous work by Stefansson et al. (2005) has shown that the H2 allele contains two copies of the proximal part of the *NSF* gene. All individuals with an H2/H2 haplotype showed MLPA ratios compatible with the presence of four copies (**Figure 5.4d**), thus further confirming our copy number estimates. In addition, all individuals with two copies as indicated by MLPA analysis were genotyped H1/H1, whereas individuals showing three copies for the *NSF* gene exhibited a H1/H2

Figure 5.2
Validation of the MLPA assay for detection of CNV exemplified by the 17q12 region and MLPA probes *TBC1D3 probe exon 4* and *TBCID3 probe exon 12* respectively. (**a**) For each region, two MLPA probes were designed. Panel A shows the correlation between the two probes, *TBC1D3 probe exon 4* and *TBC1D3 probe exon 12*, targeted to the 17q12 region in one experiment ($R^2=0.8179$). (**b**) The reproducibility was tested by determining the correlation between two successive experiments. For this, the correlation between the mean of the two *TBC1D3* probes in experiment 1 and the mean of the two *TBC1D3* probes in experiment 2 was determined ($R^2=0.7950$). (**c**) Correlation between array CGH results and MLPA results for CNV at 17q12. For this, the average array CGH $\log_2$ T/R ratios for the clones located in the CNV at 17q12 was calculated (*x* axis) and compared to the average MLPA value of the two MLPA probes for the 17q12 region (*y* axis). The diagonal black line is a regression line for the correlation between the two techniques ($R^2=0.50$). (**d**) Correlation between array CGH interpretation and MLPA results for the same 17q12 CNV locus. Based on array CGH results, patients were categorized as 'loss', 'normal' or 'gain', for showing relative copy number loss or gain, respectively (de Vries et al. 2005). On the y-axis, the average MLPA values of all patients per category are shown. The black horizontal bars indicate mean MLPA value per category for the 17q12 locus.

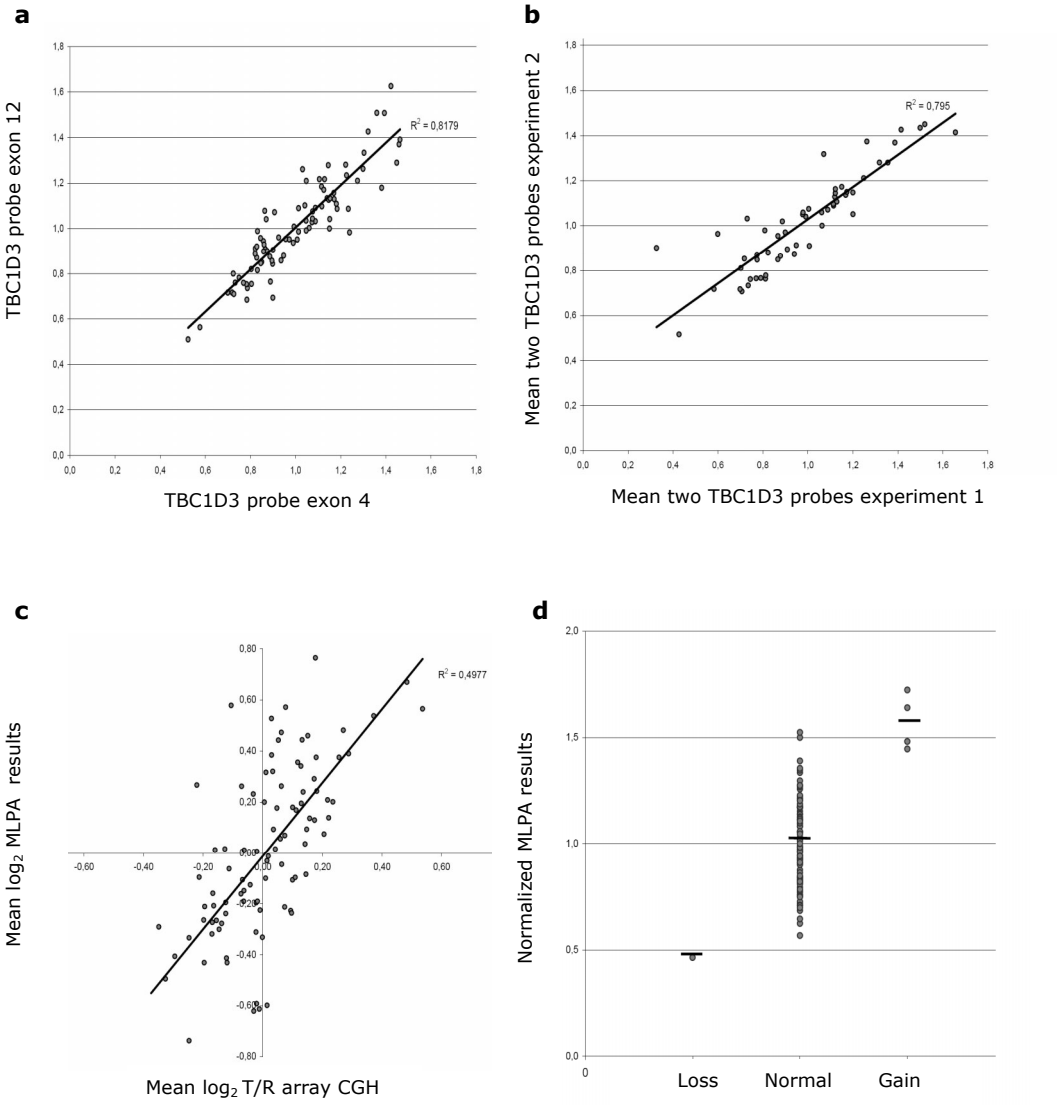

a
TBC1D3 probe exon 12
TBC1D3 probe exon 4
R² = 0,8179
b
Mean two TBC1D3 probes experiment 2
Mean two TBC1D3 probes experiment 1
R² = 0,795
c
Mean log2 MLPA results
Mean log2 T/R array CGH
R² = 0,4977
d
Normalized MLPA results
Loss
Normal
Gain

Table 5.2: Regions selected for CNV assessment in different populations

CNV region	MLPA primer location[a]	Size (kb)[b]	Gap within 100k	Overlap with LCR	Reference	Technique used[d]	Sample Size	CNV (gain/loss)	% individuals with CNV	% individuals with CNV (this study)
1p32.3	*DAB1*	119	No	No	Iafrate et al. 2004	1	55	16 (10/6)	29	0
2q31.1	*MYO3B*	161	No	No	Iafrate et al. 2004	1	55	3 (3/0)	5	0
2q35	*DES*	184	No	No	Iafrate et al. 2004	1	55	8 (4/4)	15	<1
5q13.2*	*SMN1, SMN2*	2089	No	Yes	Iafrate et al. 2004	1	55	10 (5/5)	18	7
					Sebat et al. 2004	2	20	12 (0/12)	60	
					Sharp et al. 2005	3	47	7 (6/1)	15	
					de Vries et al. 2005	4	100	13 (4/9)	13	
7p21.1	*TWIST1*	184	No	No	Iafrate et al. 2004	1	55	8 (2/6)	15	<1
7q35	*OR2F2*	838	No	Yes	Iafrate et al. 2004	1	55	7 (0/7)	13	0
					Sharp et al. 2005	3	47	9 (4/5)	19	
8p23.1*	*DEFB4, DEFB104A, DEF103B*	1410	Yes	Yes	Sebat et al. 2004	2	20	8(8/0)	40	26
					Hollox et al. 2003	5	90	17	19	
					Sharp et al. 2005	3	47	27 (6/21)	57	
					de Vries et al. 2005	4	100	5 (3/2)	5	
					McCarroll et al. 2006[c]	6	269	18	7	
					Conrad et al. 2006 [c]	6	60	3	5	
10q11.22*	*ANXA8, PPYR1, SYT15*	2569	Yes	Yes	Iafrate et al. 2004	1	55	6 (6/0)	11	7
					Sebat et al. 2004	2	20	1 (1/0)	5	
					Sharp et al. 2005	3	47	2 (2/0)	4	
					de Vries et al. 2005	4	100	4 (1/3)	4	
					McCarroll et al. 2006[c]	6	269	6	2	
					Conrad et al. 2006 [c]	6	60	3	5	

Table 5.2 (continued)

14q11*	*OR4Q3, ACTBL1*	1513	No	Yes	Sebat et al. 2004	2	20	18 (18/0)	90	11
					Bejjani et al. 2005	1	50	28 (0/28)	56	
					Sharp et al. 2005	3	47	3 (1/2)	6	
					de Vries et al. 2005	4	100	16 (5/11)	16	
17q12*	*TBC1D3*	271	Yes	Yes	de Vries et al. 2005	4	100	5 (4/1)	5	24
					Sebat et al. 2004	3	20	1 (1/0)	5	
17q21.31	*NSF*	809	No	Yes	Iafrate et al. 2004	1	55	5 (5/0)	9	52
					Sebat et al. 2004	2	20	3 (3/0)	15	
					Sharp et al. 2005	3	47	15 (10/5)	32	
22q11.21*	*PRODH, GGT2*	454	No	Yes	Sharp et al. 2005	4	47	24 (0/21)	51	3
					de Vries et al. 2005	4	100	12 (4/8)	12	

[a]: MLPA primer sequences are listed in **Table 5.1**
[b]: Maximum genomic size based on overlapping CNVs reported
[c]: Detection of deletion variants only
[d]: Technique codes (1) 1 Mb BAC array; (2) ROMA; (3) Targeted CNV array; (4) 100 kb BAC array; (5) MAPH; (6) SNP array
*: Region validated before assessing CNV in different populations

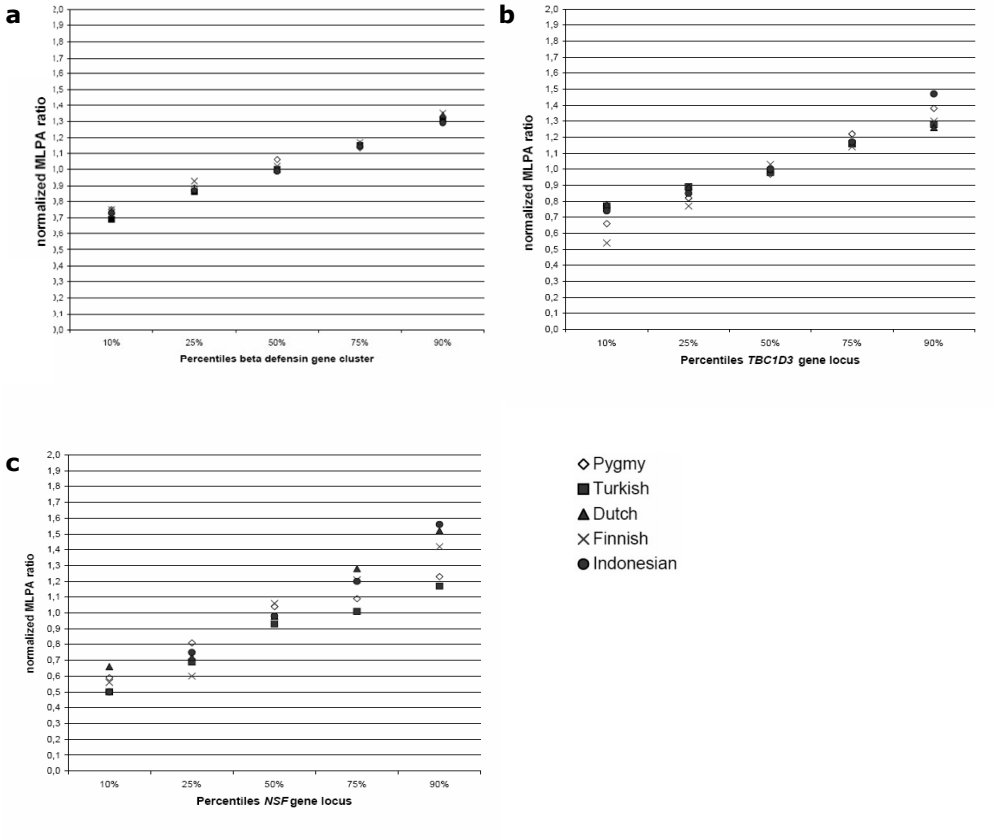


Figure 5.3
Interpopulation percentile differences for the beta defensin gene cluster and the *TBC1D3* and *NSF* gene loci. MLPA normalized ratios at the percentiles 10%, 25%, 50% 75% and 90% for the beta defensin gene cluster (**a**), *TBC1D3* gene locus (**b**) and *NSF* gene locus (**c**). For the beta defensin gene cluster, the percentiles differ <10% between populations, whereas for the *TBC1D3* and *NSF* gene loci differences were observed that suggest distinct inter-population differences in copy number distribution.

heterozygous haplotype (**Figure 5.4d**). Together, these results suggest a relationship between the haplotype and the *NSF* copy numbers observed.

Discussion

In contrast to SNPs, of which over one million variants and their frequencies in four different ethnic populations have recently been described (International HapMap consortium 2005), little is known about the relative DNA copy numbers and frequencies of CNVs in different ethnic populations. Previous studies, using a combination of different array platforms and computational algorithms, have provided a first insight into the

Table 5.3: CNV frequency in ethnic populations

CNV region	**1p32.3**		**2q31.1**		**2q35**		**5q13.2**		**7p21.1**		**7q35**		**8p23.1**	
Population	loss[a]	Gain[b]	loss	gain	loss	gain	loss	gain	loss	gain	loss	gain	loss	gain
Pygmy *(n=37)*	0	0	0	0	0	0	5	5	0	0	0	0	11	21
Indonesian *(n=88)*	0	0	0	0	0	0	2	2	1	0	0	0	14	13
Dutch *(n=84)*	0	0	0	0	0	0	2	6	0	0	0	0	10	15
Turkish *(n=64)*	0	0	0	0	nd	nd	2	6	nd	nd	0	0	13	16
Finnish *(n=36)*	0	0	0	0	6	0	3	0	0	0	0	0	6	17
Total (n=309)	0	0	0	0	<1	0	3	4	<1	0	0	0	11	15

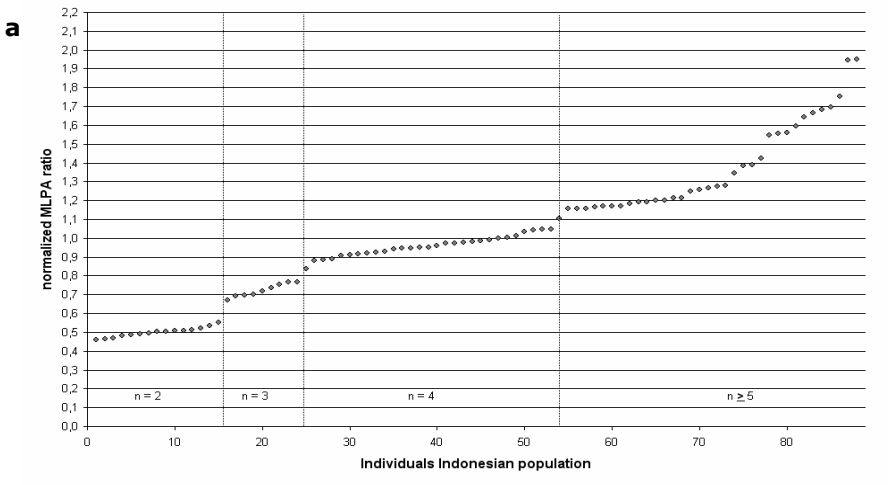
CNV region	**10q11.22**		**14q11**		**17q12**		**17q21.31-*NSF* exon 2,9,12**		**17q21.31-*NSF* exon 14[c]**		**22q11.21**	
Population	loss	gain	loss	gain	loss	gain	loss	gain	loss	gain	loss	gain
Pygmy *(n=37)*	0	8	0	5	16	16	19	11	0	0	3	5
Indonesian *(n=88)*	0	3	2	6	11	19	24	20	1	1	1	1
Dutch *(n=84)*	1	6	4	6	4	10	28	29	0	0	0	1
Turkish *(n=64)*	3	8	8	9	8	13	23	36	0	0	0	2
Finnish *(n=36)*	0	11	3	6	17	11	31	25	3	0	6	6
Total (n=309)	1	6	4	7	10	14	24	28	<1	<1	1	2

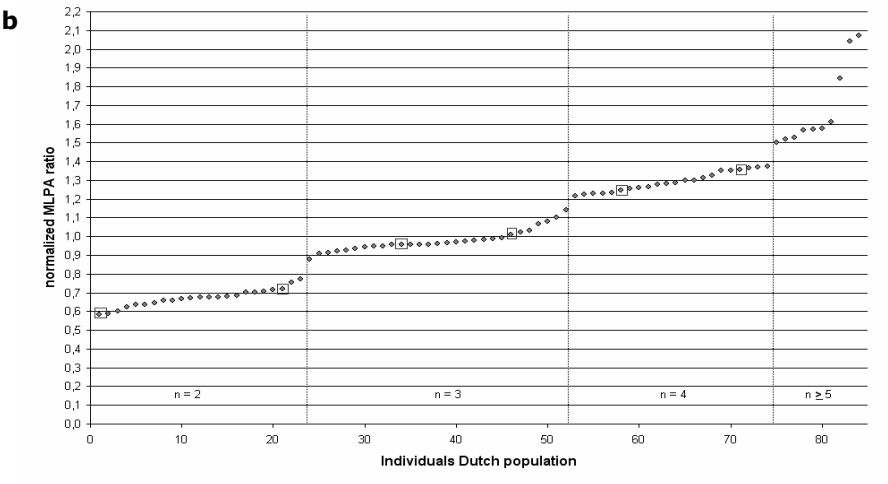
[a]: Percentage of individuals with at least two probes having a normalized ratio below 0.75
[b]: Percentage of individuals with at least two probes with a normalized ratio above 1.25
[c]: Calculations for *NSF* exon 14 are based on a single probe. nd = not determined

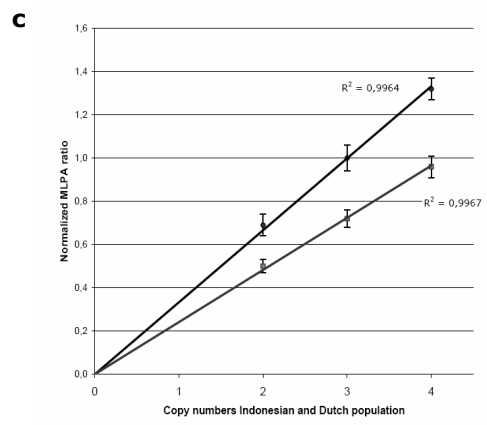
occurrence of CNVs in normal healthy individuals (Fredman et al. 2004; Iafrate et al. 2004; Sebat et al. 2004; de Vries et al. 2005; Sharp et al. 2005; Tuzun et al. 2005; Conrad et al. 2006; Hinds et al. 2006; McCarroll et al. 2006). Here, we have focused on twelve genomic regions previously identified as showing recurrent CNV. Using multiplex ligation-dependent probe amplification (MLPA), a technique perfectly suited for semi-quantitative analysis of a large cohort of individuals (Schouten et al. 2002; Koolen et al. 2004; Rooms et al. 2006), we found significant CNV in seven of these twelve regions in all ethnic populations studied. In addition, we observed significant inter-population differences in DNA copy number frequencies for the *TBC1D3* and *NSF* gene loci.

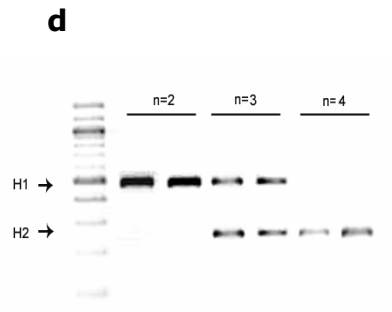
The five regions for which no CNV was detected differ from the other seven in several aspects. Four of these regions were reported to show CNV in only one study (Iafrate et al. 2004), whereas all other regions were reported to show CNV in multiple studies. Related to this, the identification of these five CNVs in the original study by Iafrate et al. (2004) was based on a single array target, i.e., one BAC clone, whereas the other seven CNV regions were identified by multiple array targets. The fact that these results were based on only one measurement in one study, not validated by other technologies, questions the validity of these initial CNV observations. The genomic architecture of these five regions also differs from the other seven. All regions for which CNV was identified in this study were flanked or (partly) covered by LCRs. Additional indications for the involvement of LCRs in these common CNV regions are provided by our detailed analysis of the NSF gene locus for which four MLPA probes were tested. Three of these probes were located within a LCR that partly covers the NSF gene, whereas one probe was located more proximally, outside this LCR. CNV was only observed for the MLPA probes that were located within the LCR. Our data therefore support the hypothesis that recurrent large-scale and intermediate-sized CNVs ('CNV hotspots') are mediated by LCRs. In addition, there may be a second class of CNVs that is only detected in single parent-offspring trios ('rare variants'), which may not be mediated by specific genomic architectural features such as LCRs. The five regions not showing CNV in this study could represent such rare variants, although they were specifically chosen based on presumed frequencies of >5% reported in one or more

Figure 5.4
Examples of normalized copy number ratios of the *NSF* gene for the Indonesian and Dutch populations. For each individual the average value of the probes within the CNV was calculated and, subsequently, ordered from low to high. Subgroups corresponding to different DNA copy numbers were defined by eye and cluster analysis, and are delineated by vertical dotted lines. By calculating the relative differences between each subgroup ('steps'), we estimated the actual DNA copy numbers. For the Indonesian population (**a**) the lowest ratio is ~0.50 with steps of ~0.25. For the Dutch population (**b**) the lowest ratio is ~0.66 with steps of ~0.33. The lowest copy number in each population is two copies, indicating that the median copy number within the Indonesian population is four whereas for the Dutch population a median copy number of three is observed. Both populations, however, show a similar copy number range from two to at least seven copies. The best-fit regression analyses (**c**) show that copy numbers are reliably estimated for the Indonesian (R^2 =0.9967; red) and Dutch populations (R^2=0.9964; black), respectively. (**d**) Genotypes for H1 and H2 of individuals of the Dutch population (indicated by red boxes in panel **c**). All individuals exhibiting the presence of two copies by MLPA analysis were genotyped H1/H1, whereas individuals with three copies showed a H1/H2 genotype. In addition, a number of the individuals showing four copies by MLPA analysis showed a homozygous H2/H2 genotype.


a
normalized MLPA ratio
n = 2
n = 3
n = 4
n ≥ 5
Individuals Indonesian population


b
normalized MLPA ratio
n = 2
n = 3
n = 4
n ≥ 5
Individuals Dutch population


c
Normalized MLPA ratio
R² = 0,9964
R² = 0,9967
Copy numbers Indonesian and Dutch population


d
n=2
n=3
n=4
H1
H2

previous studies. Several examples of this class of rare variants have been reported before (de Vries et al. 2005; Conrad et al. 2006). Alternatively, technical aspects may be the explanation for not detecting CNV for these five regions. One such aspect is the MLPA probe location. For each region, the MLPA probes are targeted to only (a few) hundred basepairs of the BAC clones described. Therefore, if only a part of the BAC clone (with an average genomic size of ~170 kb) shows CNV, this CNV may be missed by the MLPA assay.

For seven genomic regions, significant CNV was detected in all populations. We analyzed this variation by two different approaches. First we used a standard classification into copy number gain, copy number loss or no copy number variation, using MLPA thresholds for copy number gain and loss of 1.25 and 0.75, respectively. By doing so, copy number variation of these seven loci was observed in all populations tested. Furthermore, we found that the percentages of the population showing CNV for a locus, i.e., the percentages of individuals with copy numbers other than the population median, were virtually identical between the populations tested. Regions for which most variation was observed (24-52%) included 8p23.1, 17q12 and 17q21.31. For four other regions, 5q13.2, 10q11.22, 14q11 and 22q11.21, we observed CNV at lower frequencies (3-11%). The percentages of individuals showing CNV for these seven regions were within the range of frequencies as reported previously by others (see **Table 5.2**). Although this analysis was useful for gaining insight into the frequencies of occurrence, it failed to provide detailed information on the actual copy numbers of the CNV loci tested.

Therefore, as a second approach, we set out to obtain such detailed information from our semi-quantitative MLPA analyses. After ranking the MLPA values from low to high per population, we could clearly define subgroups for (part of) the *NSF* gene. These MLPA-defined subgroups indicated a copy number range from two to seven in four out of five populations tested, which is in concordance with a previous report by Stefansson et al. (2005). The frequencies at which these actual copy numbers were detected within populations differed markedly. In addition, the population-specific median copy numbers varied between populations. By using traditional approaches for looking at CNVs, these inter-population differences would have escaped attention. In fact, based on our first approach, these populations seemed rather similar to each other regarding CNV, all having ~50% of individuals showing CNV (see **Table 5.3**). Furthermore, this semi-quantitative copy number analysis resulted in higher CNV percentages as compared to the standard copy number analysis. For instance, after estimation of the actual copy numbers for the *NSF* gene locus, approximately 70% of individuals within four of the populations tested showed copy numbers different from the population median, as compared to 52% of the individuals by our first approach. These observations strongly suggest that the standard approach of classifying individuals into 'loss', 'normal' and 'gain' to characterize CNVs is inadequate.

The 17q21.31 region, containing the *NSF* gene, is associated with a complex genomic architecture, involving an inversion polymorphism (Stefansson et al. 2005). Two main haplotypes have been identified for this inversion, designated H1 and H2, respectively. The H2 lineage, representing the inversion polymorphism, is found at a frequency of 20% in European populations, whereas a frequency of 1% is found in Asian populations. Interestingly, *de novo* deletions within this inversion polymorphism have recently been

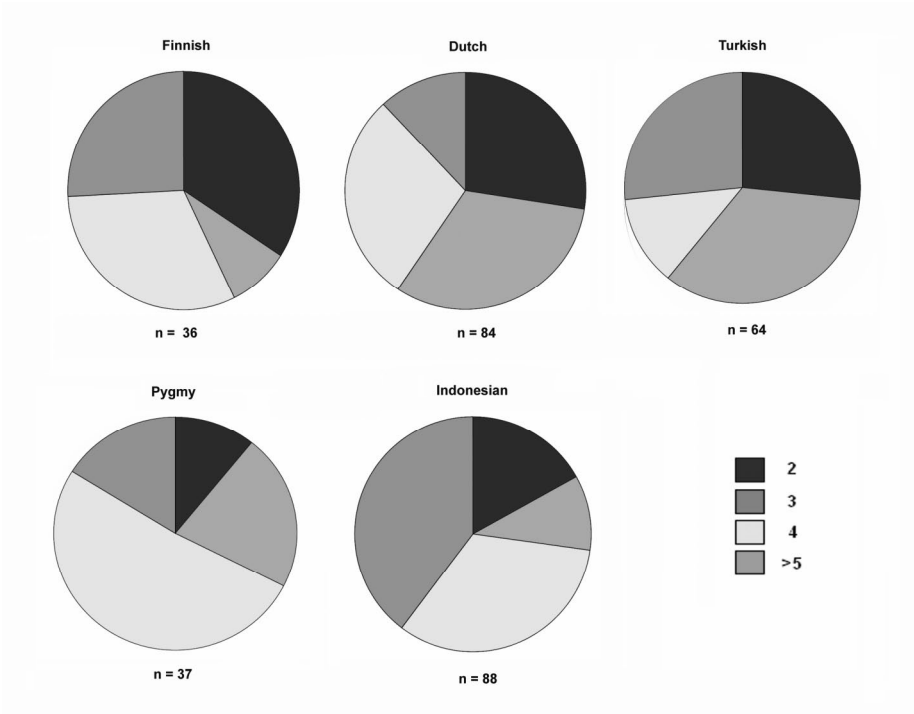


Figure 5.5
Inter-and intra-population differences in *NSF* copy number estimations. Different copy numbers are displayed as a percentage of the total population (see also **Figure 5.4**). The most significant difference is seen between the Dutch and Indonesian populations. ($p<0.0001$).

described and define a previously unrecognized genomic disorder, potentially accounting for as many as 1% of all mental retardation cases (Chapter 4; Shaw-Smith et al. 2006; Sharp et al. 2006). Our results showed that Dutch individuals homozygous for the H2 haplotype carry four copies of the *NSF* region exhibiting CNV (exons 2, 9 and 12), whereas individuals with two copies showed a H1/H1 haplotype. These results indicate that the H1 allele contains one *NSF* copy whereas the H2 allele contains two *NSF* copies, which is in concordance with previous results (Stefansson et al. 2005). This notion is further substantiated by the observation that H1/H2 heterozygous individuals showed three copies of the *NSF* gene. We observed a small percentage of Indonesian individuals with three *NSF* copies, which may be explained by the low frequency of the H2 haplotype in Asian populations. The presence of copy numbers of four and higher can be explained by the fact that the H1 haplotype can be subdivided into H_{D1} and H_{D3}, containing one and three copies of the proximal part of the *NSF* gene, respectively (Stefansson et al. 2005). This subdivision may also explain why not all individuals of the Dutch population with four copies showed a homozygous H2/H2 genotype, i.e., individuals homozygous for H1 but with H_{D1} and H_{D3} alleles will also have four copies of the proximal part of the *NSF* gene.

Another region for which actual copy numbers were estimated is the beta defensin gene cluster region at 8p23.1. Assuming that the lowest copy number is two, the majority of individuals of all populations have either three or four copies of the beta defensin gene, with a copy number range from two to seven, which is consistent with previous studies (Hollox et al. 2003). In contrast to the *NSF* locus, we did not observe any differences in distribution of the actual copy numbers between populations. Estimation of the actual copy numbers appeared to be more problematic for the *TBC1D3* gene (17q12), because no obvious subgroups were observed. These results are consistent with relatively high copy numbers in all populations tested. Further support for this observation comes from the signal intensity of the two *TBC1D3* probes used, which was consistently the highest of all probes present in the assay. In agreement with this notion, detailed analysis of the LCR structures within this region showed that *TBC1D3*, encompassing the MLPA probes, is located within an LCR which is present in eight copies in 17q12.

The first biological consequences of copy number variation for a single gene have recently been reported (Gonzalez et al. 2005; Aitman et al. 2006). Here, we show the presence of a continuous spectrum of copy numbers in several CNV hotspots, ranging from two to at least seven. We and others hypothesize that different copies within this spectrum may have different phenotypic consequences and, ultimately, may lead to disease if copy numbers exceed certain threshold levels (Pinkel & Albertson 2005). Similar mechanisms are known to cause monogenic diseases such as autosomal dominant facioscapulohumeral muscular dystrophy (FSHD1A), which is known to be due to variations in the D4Z4 repeat. This repeat consists of identical 3.3 kb units, varying from 11 to 100 copies (33 kb – ~300 kb) in normal healthy individuals, whereas patients with FSHD1A carry 1–10 copies (Wijmenga et al. 1992; van Deutekom et al. 1993). We suggest that insight into the actual DNA copy numbers for loci exhibiting CNV is required to determine their potential contribution to normal phenotypic variation and, in addition, disease susceptibility.

In conclusion, we have shown CNV for seven out of twelve selected loci in multiple populations using MLPA, and analyzed the observed variation by two approaches. A detailed analysis of actual copy numbers showed inter- and intra-population differences, implying that there is more to studying CNV than 'simply' scoring individuals for copy number gain or loss. Such information may turn out to be essential for understanding the role of DNA copy numbers in normal phenotypic variation and disease susceptibility.

Acknowledgements

The authors thank Rolph Pfundt and Bert de Vries for useful discussions, patient collection and initial array CGH analyses, and the Leiden Genome Technology Center (LGTC) for technical support. This work was supported by grants from the Netherlands Organization for Health Research and Development (MHB/HB/JAV, ZonMW 912-04-047, JAV ZonMW 917-66-363).

Chapter 6

Complex chromosome 17p rearrangements associated with low-copy repeats in two patients with congenital anomalies

Lisenka ELM Vissers, Paweł Stankiewicz, Svetlana A Yatsenko, Eric Crawford, Heather Creswick, Virginia K Proud, Bert BA de Vries, Rolph Pfundt, Carlo LM Marcelis, Joleen Zackowski, Weimin Bi, Ad Geurts van Kessel, James R Lupski & Joris A Veltman

Human Genetics 121: 697-709, 2007

Abstract

Recent molecular cytogenetic data have shown that the constitution of complex chromosome rearrangements (CCRs) may be more complicated than previously thought. The complicated nature of these rearrangements challenges the accurate delineation of the chromosomal breakpoints and mechanisms involved. Here, we report a molecular cytogenetic analysis of two patients with congenital anomalies and unbalanced *de novo* CCRs involving chromosome 17p using high-resolution array-based comparative genomic hybridization (array CGH) and fluorescent *in situ* hybridization (FISH). In the first patient, a 4-month-old boy with developmental delay, hypotonia, growth retardation, coronal synostosis, mild hypertelorism, and bilateral club feet, we found a duplication of the Charcot-Marie-Tooth disease type 1A (CMT1A) and Smith-Magenis syndrome (SMS) chromosome regions, inverted insertion of the Miller-Dieker lissencephaly syndrome (MDLS) region into the SMS region, and two microdeletions including a terminal deletion of 17p. The latter, together with a duplication of 21q22.3-qter detected by array CGH, are likely the unbalanced product of a translocation t(17;21)(p13.3;q22.3). In the second patient, an 8-year-old girl with mental retardation, short stature, microcephaly and mild dysmorphic features, we identified four submicroscopic interspersed 17p duplications. All 17 breakpoints were examined in detail by FISH analysis. We found that four of the breakpoints mapped within known low-copy repeats (LCRs), including LCR17pA, middle SMS-REP/LCR17pB block, and LCR17pC. Our findings suggest that the LCR burden in proximal 17p may have stimulated the formation of these CCRs and, thus, that genome architectural features such as LCRs may have been instrumental in the generation of these CCRs.

Introduction

The majority of constitutional chromosomal abnormalities are thought to be simple rearrangements, involving less than three breaks in one or two chromosomes. It is known that genomic architectural features such as low-copy repeat (LCR) structures may play an important role in the formation of these abnormalities (Lupski 1998; Stankiewicz & Lupski 2002). LCRs, also termed segmental duplications or duplicons, encompass apparently normal stretches of genomic DNA, often containing genes. LCRs are present in more than one copy in the genome and are defined by a ≥ 1 kb size and a >90% sequence identity (Bailey et al. 2001; Eichler 2001; Bailey et al. 2002). LCRs can mediate recurrent DNA rearrangements such as deletions, duplications and inversions through chromosome or chromatid misalignment followed by nonallelic homologous recombination (NAHR). The majority of the currently known chromosome microdeletion/duplication syndromes result from NAHR between large (usually >10 kb), highly identical (>95%) LCRs and are, therefore, also termed genomic disorders (Lupski 1998; Stankiewicz & Lupski 2002; Lupski & Stankiewicz 2006). In addition to recurrent events, breakpoints of nonrecurrent rearrangements have been associated also with genomic architectural features including LCRs and AT-rich palindromic sequences (Stankiewicz et al. 2003; Shaw & Lupski 2004; Shaw & Lupski 2005; Kriek et al. 2006; Lee et al. 2006).

Complex chromosome rearrangements (CCRs) are relatively rare events, and can be classified into groups based on the number of breakpoints and type of rearrangement,

being (i) three way exchange with three breaks in three involved chromosomes, (ii) two way exchange that coincide with two separate reciprocal translocations and (iii) exceptional CCRs with multiple breaks and complicated rearrangements (Bartels et al. 2007). CCRs have been observed in phenotypically normal individuals as well as in individuals with mental retardation and/or congenital abnormalities (Pai et al. 1980; Battisti et al. 2003; Kuechler et al. 2005). As a consequence, the characterization of these rearrangements is essential for reaching a proper clinical diagnosis and for estimation of the recurrence risk. In standard clinical practice, this characterization is usually achieved by conventional cytogenetic approaches only, thus submicroscopic imbalances at the breakpoints will remain undetected because of the low resolution of such approaches. Recently, the use of high-resolution molecular techniques such as array-based comparative genomic hybridization (array CGH) have contributed to a growing awareness of the presence of CCRs and cryptic imbalances in patients with MR and/or congenital anomalies (Vissers et al. 2003; Weise et al. 2003; Lespinasse et al. 2004; Patsalis et al. 2004; Shaw-Smith et al. 2004; Vermeulen et al. 2004; Borg et al. 2005; de Vries et al. 2005; Chen et al. 2006; Gajecka et al. 2006; Karmous-Benailly et al. 2006). Although molecular mechanisms have been studied in recurrent, simple rearrangements, the role of genomic architecture underlying the occurrence of nonrecurrent CCRs, remains as yet poorly understood because until recently detailed identification of the exact breakpoints was lacking.

Here, we investigated the potential involvement of genome architectural features such as LCRs in the occurrence of nonrecurrent 17p CCRs in two patients with congenital anomalies. The proximal chromosome 17p arm is associated with a wide variety of recurrent chromosome aberrations resulting from NAHR between LCRs including four well known genomic disorders: Charcot-Marie-Tooth disease type 1A (CMT1A) [MIM 118220], hereditary neuropathy with liabilities to pressure palsies (HNPP) [MIM 162500], Smith-Magenis syndrome (SMS) [MIM 182290] and the Potocki-Lupski syndrome associated with duplication 17p11.2 (Pentao et al. 1992; Reiter et al. 1996; Chen et al. 1997; Potocki et al. 2000; Inoue et al. 2001; Shaw et al. 2002; Bi et al. 2003; Potocki et al. 2007). By applying high-resolution array CGH to these two novel CCRs we observed an unexpected level of complexity. Subsequent FISH analyses revealed the presence of LCRs at the CCR breakpoints, thus providing a potential mechanistic basis for the occurrence of the CCRs studied. We propose that genomic architectural features such as LCRs may underlie the occurrence of these CCRs.

Patients, Materials, and Methods

Clinical description of patients

Patient 1

The male proband was born to a 42-year-old woman and 43-year-old man at 39 week's gestation by Caesarean section. Because of an abnormal ultrasound showing prominent nuchal fold, growth retardation, club feet, and hyperechoic bowel, amniocentesis was performed at 20 weeks of gestation and an abnormal karyotype with 17pter monosomy and 17p13.1p13.3 trisomy was identified. The birth weight was 2,840 g (10^{th} centile), length 51 cm (60^{th} centile), and head circumference 36 cm (50^{th} centile). Apgar scores were 6 and 8 at 1 and 5 minutes, respectively. The bilateral club feet were corrected by

surgery. Echocardiogram revealed bicuspid aortic valve. Physical exam at 11 months revealed weight and height at the 3rd centile while his head circumference had increased to the 75th centile. There was a premature closure of the metopic suture, with anterior fontanel still open (2 × 2 cm) and mild hypertelorism. At the age of 16 months, he was an amiable, interpersonally interactive little boy who had developmental delay with babbling but no clear words. In addition, he suffered from truncal hypotonia but he was able to pull to stand. His skin was dry with marked eczema and the right testis was still undescended.

Patient 2

This 8-year-old, moderately mentally retarded girl was born after an uneventful pregnancy at 40 week's gestation, with a normal birth weight of 3,400 g. Apgar scores were 3, 5, and 6 at 1 min, 5 min, and 10 min, respectively. Feeding problems were present from birth. Her development was moderately delayed with an IQ score of 50 points at the age of 4 years and 9 months. At the age of 6 years and 10 months, she had a short stature of 110 cm (< 3rd centile) and a head circumference of 49 cm (10th centile) with distinctive facial features such as synophrys, upward slanting palpebral fissures, flattened upper-ear helices with large ear lobules, an upturned nasal tip, and a broad mouth with a long philtrum and full lower lip. She had a cardiac murmur; short broad hands with clinodactyly of the fifth fingers and absence of distal interphalangeal joint creases; short broad feet with short toes and a slight skin syndactyly between the second and third toes bilaterally. She had an outgoing personality and easily approaches strangers. At the age of 8 years, she started to complain about painful feet and developed a bilateral pes cavus deformity. Neurological examination showed areflexia and electromyography (EMG) studies were consistent with a demyelinating neuropathy. In addition, treatment for precocious puberty was started.

We obtained samples from the patients and their family members after acquiring informed consent approved by the Institutional Review Board for Human Subject Research at Baylor College of Medicine, Houston, and Radboud University University Nijmegen Medical Centre, Nijmegen.

Array-based comparative genomic hybridization

Genomic DNA was isolated from peripheral blood lymphocytes or from lymphoblastoid cell lines by routine procedures. Two pools of reference DNA were used – one containing equal amounts of genomic DNA from 10 healthy male blood donors and one containing equal amounts of genomic DNA from 10 healthy female blood donors. Isolation of genomic DNA, DNA labeling, hybridization of labeled DNA to a 32,447-BAC array, and spot identification were performed as described elsewhere (de Vries et al. 2005). In brief, 500 ng of genomic DNA from each patient was labeled by random priming with Cy3-dUTP or Cy5-dUTP (Amersham Biosciences) and hybridized to the sex-mismatched reference pool. Test and reference samples were mixed with 120 µg of human Cot-1 DNA (Roche), co-precipitated, and resuspended in 120 µl of a hybridization solution containing 50% formamide, 10% dextran sulfate, 2×SSC, 4% SDS, and 10 mg/ml of yeast tRNA (Invitrogen). Hybridization and post-hybridization washing procedures were performed using a GeneTac Hybridization Station (Genomic Solutions). An 18-h hybridization at 37°C with active circulation of the probe was performed, followed by five post-hybridization wash cycles in 50% formamide and 2×SSC at 45°C, and five wash cycles in phosphate buffer at 20°C. Slides were dried by centrifugation and scanned using a GenePix Autoloader 4200AL laser scanner (Axon Instruments). Spot dentification and two-color fluorescence intensity measurements were obtained using the GenePix 5.1 software, and all data were entered into a database for

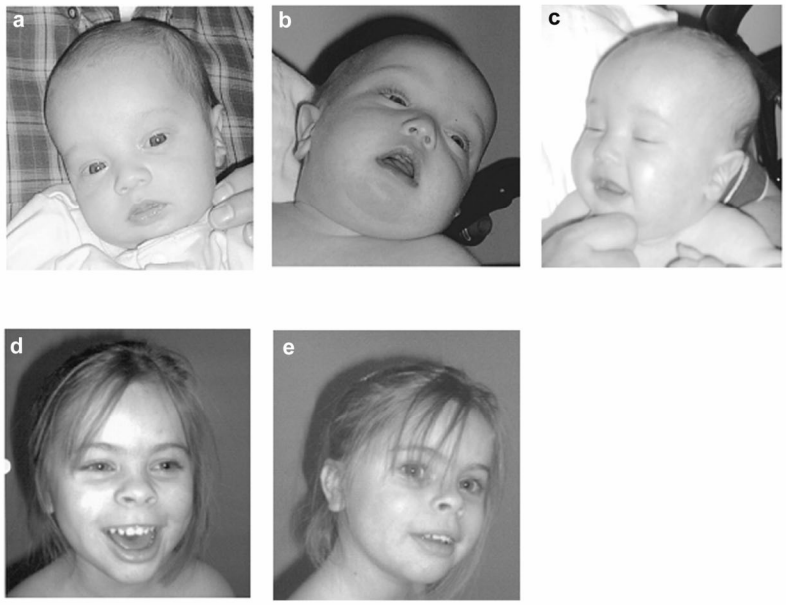


Figure 6.1.
Patient 1 (**a-c**) at the age of 4 months. Patient 2 (**d-e**) at the age of 6 years and 10 months.

subsequent analysis. Following normalization, the $\log_2$-transformed test-over-reference ratios were analyzed for loss and gain of genomic regions by a standard Hidden Markov Model (HMM) (Rabiner 1989; de Vries et al. 2005).

Fluorescent in situ hybridization (FISH)

PAC and BAC probes specific for human chromosome 17p were selected based on their physical location within the affected 17p region (Inoue et al. 2001; Bi et al. 2002; http://genome.ucsc.edu/). DNA was isolated from liquid cultures using Perfectprep Plasmid Mini (Eppendorf, Hamburg, Germany). The relative alignments of the selected BACs were determined by BLAST searches against the high-throughput genome sequence database (http://www.ncbi.nlm.nih.gov/blast) and assembled using the Sequencher software (Gene Codes Corp., Ann Arbor, MI, USA). FISH was essentially performed as described (Stankiewicz et al. 2001a). In brief, 100–200 ng of isolated BAC or PAC DNA was labeled with biotin or digoxigenin by nick-translation (BioNick Labeling System, Invitrogen; DIG-Nick Translation Mix, Roche) and visualized with FITC avidin (Vector) or rhodamine-labeled antibodies (Sigma). Patient's chromosomes derived from lymphoblast cell lines were counterstained with DAPI (Sigma). The signals from the normal chromosomes 17 were used as internal control. Cells were analyzed using a Zeiss Axioskop fluorescence microscope equipped with an appropriate filter combination and a CCD camera. Monochromatic images were captured and pseudocolored using MacProbe 4.2.2 (Perceptive Scientific Instruments, League City, TX, USA) on a Power Macintosh G4 system.

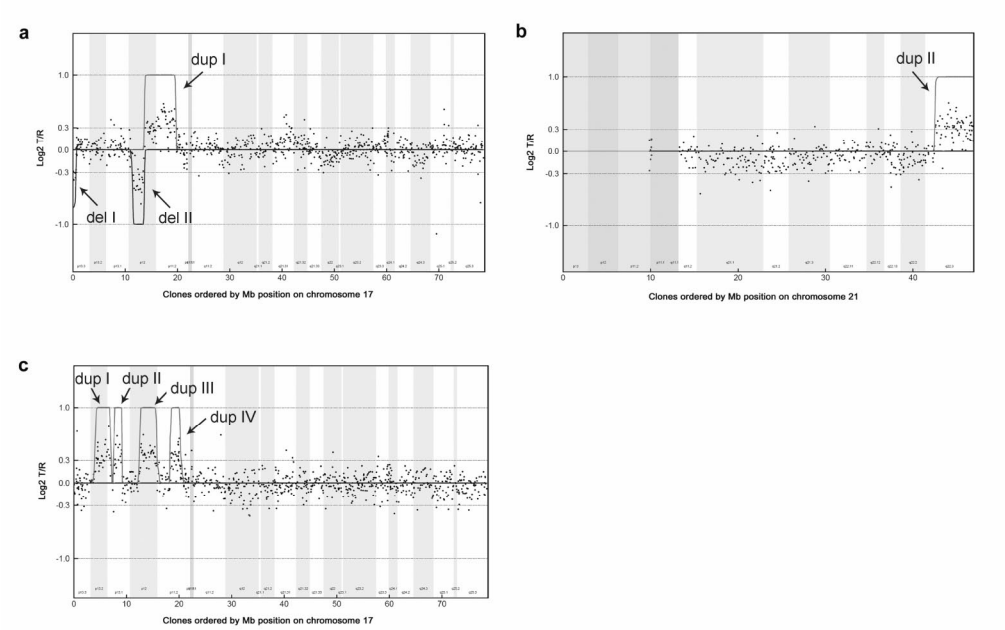


Figure 6.2
Chromosome 17 (patient 1 and 2) and chromosome 21 (patient 1) array CGH profiles. On the *x* axis, clones are ordered by Mb position on chromosome 17 and 21, respectively, and on the *y* axis $\log_2$ T/R ratios are shown. Hidden Markov Model was used to identify the duplications (green lines) and deletions (red lines) in patient 1 and 2. (**a**) In patient 1, chromosome 17 shows two interspersed deletions of ~600 kb (17p13.3 – Del I) and ~2.2 Mb (17p12 – Del II), respectively, and a duplication of ~6.1 Mb (Dup I). (**b**) Additionally, a 4.4 Mb duplication of 21q22.3 was observed (Dup II). (**c**) Patient 2 showed four interspersed duplications on 17p11-p13 (Dup I - Dup IV), in total comprising ~8.8 Mb of genomic sequence. All alterations were proven to be *de novo*.

Computational analyses

For the identification of novel LCRs associated with the CCRs, ~100 kb fragments of genomic sequence flanking breakpoints for which no LCRs are known currently, were analyzed for sequence homology (http://genome.ucsc.edu; May 2004 freeze). Interspersed repeat sequences within the downloaded DNA sequence were eliminated by RepeatMasker (http://genome.ucsc.edu) and the repeat masked genomic sequences were analyzed using NCBI BLAST2 (http://www.ncbi.nlm.nih.gov/blast/ bl2seq/bl2.html).

Results

Cytogenetic and molecular analyses of patient 1

Conventional cytogenetic analysis at 550-band resolution of patient 1 (**Figure 6.1a-c**) revealed an abnormal male karyotype with an apparent complex rearrangement involving the short arm of chromosome 17. Parental chromosomes at 600-band resolution were normal. FISH analysis confirmed this complex rearrangement and, in addition, revealed a

Table 6.1: Summary of the array CGH and FISH results patient 1

Chr. band	Clone[a]		Distance from pter (Mb)[e]	Der (17)	Technique(s)	
17p13.3	CTD-2348K1		0.1	-	array CGH	
17p13.3	CTD-2326F1	*R*	0.1	-	array CGH/ FISH	
17p13.3	CTD-2573J8		0.2	-	array CGH	del I
17p13.3	RP11-1260E13	*G*	0.3	-	FISH	
17p13.3	RP11-143G11		0.6	-	array CGH	
17p13.3	RP11-488O17		0.6	normal	array CGH	
17p13.3	RP11-411G7		0.6	normal	FISH	
17p13.3	RP1-95H6[b]	*R*	2.7	inv ins	FISH	
17p13.2	RP11-148L19		4.1	inv ins	FISH	***middle SMS-REP/ LCR17pB***
17p13.2	GS-202L17[c]	*G*	4.9	inv ins	FISH	
17p13.2	RP11-373N8		5.5	inv ins	FISH	
17p12	RP11-466J1		11.1	normal	array CGH	
17p12	RP11-22L4		11.2	normal	FISH	
17p12	RP11-271N1		11.4	normal	array CGH	
17p12	RP11-165H21		11.5	-	array CGH	
17p12	RP11-802E21		11.6	-	array CGH	
17p12	RP11-64L11		13.1	-	FISH	del II
17p12	RP11-352N24		13.2	-	array CGH	
17p12	RP11-590H8		13.4	-	array CGH	
17p12	RP11-112H7		13.6	normal	array CGH	
17p12	CTD-2120F7		13.7	normal	array CGH	
17p12	RP11-687M21		13.9	++	array CGH	
17p12	RP11-78J16		14.1	++	array CGH	
17p12	RP1-150M12[d]	*R*	15.4	++	FISH	
17p11.2	RP11-209J20		16.3	++	FISH	
17p11.2	RP11-525O11	*G*	17.6	++	FISH	dup I
17p11.2	RP1-178F10		18.4	++	FISH	
17p11.2	RP11-28B23		18.7	++	FISH	
17p11.2	RP11-1113L8		19.3	++	FISH	
17p11.2	RP11-277B5		19.4	++	FISH	
17p11.2	CTD-2019P4		19.6	++	array CGH/ FISH	
17p11.2	CTD-2020J20		19.7	normal	array CGH	
17p11.2	RP11-78O07		19.7	normal	FISH	

Table 6.1 (continued)

Chr. band	Clone[a]		Distance from pter (Mb)[e]	Der (21)	Technique(s)	
21q22.3	RP11-690D9		42.5	normal	array CGH	
21q22.3	RP11-282I20		42.6	++	array CGH	
21q22.3	RP11-40L10	*G*	44.2	++	FISH	dup II Located der(17)
21q22.3	RP11-16B19	*R*	45.1	++	FISH	
21q22.3	RP11-1000I21		46.8	++	array CGH	

'-' Deleted; '++' duplicated; 'inv ins' inverted inversion. Extent of deletion and duplication is indicated at the right; Breakpoint-associated LCRs are indicated in bold.
[a]: Clones indicated by *R* (for red) or *G* (for green) are shown in **Figure 6.3a-d** in the respective colors
[b]: *LIS1*-specific PAC clone (Chong et al. 1997)
[c]: Adapted from Knight et al. 2000
[d]: *PMP22*-specific PAC clone
[e]: Distance from pter to the clone midpoints

deletion of 17pter, a duplication of the genomic region containing the SMS and the CMT1A regions, and an inversion involving the Miller-Dieker lissencephaly syndrome (MDLS) region. The presence of these abnormalities in the parental chromosomes was excluded by FISH. The patient's karyotype was designated as:

46,XY,der(17).ish del(17)(p13.3)dup(17)(p11.2p12)inv(17)(p11.2p13.3) dn.

Subsequently, array CGH was used to refine the breakpoints and genomic sizes of the complex 17p rearrangement. The deletion involving the telomeric region of 17p was determined to be ~600 kb in size, whereas the duplication containing the SMS and CMT1A region was ~6.1 Mb in size. Interestingly, array CGH revealed two additional submicroscopic alterations that had remained undetected using previous approaches. These included an ~2.2 Mb interstitial deletion in 17p12 and an ~4.4 Mb duplication in 21q22.3 (**Figure 6.2a-b**).

Dual color FISH experiments were performed to independently map each of the chromosome breakpoints (**Figure 6.3; Table 6.1**). FISH analyses confirmed the deletion of the 17p telomeric region (**Figure 6.3a**) as well as the dup(17)(p11.2p12) (**Figure 6.3b**). The FISH results of the duplication were consistent with a tandem duplication (**Figure 6.3b**). Interestingly, the MDLS region was found to be inserted into the middle SMS-REP/LCR17pB block in the genomic region involved in SMS (**Figure 6.3c**). Unexpectedly, the duplicated fragment 21q22.3qter was localized on der(17), likely representing the product of an unbalanced translocation t(17;21)(p13.3;q22.3) (**Figure 6.3d**). FISH analyses on the parental chromosomes excluded the presence of a balanced t(17;21)(p13.3;q22.3). Based on these results, the patient's karyotype was redefined as:

46,XY,der(17).ish del(17)(p13.3)del(17)(p12p12)dup(17)(p11.2p12)ins(17)(p11.2p13.3p13.2)t(17;21)(p13.3;q22.3). arr cgh 17p13.3(CTD-2348K1-RP11-143G11)x1, 17p12(RP11-165H21-RP11-590H8)x1,17p12p11.2(RP11-687M21-CTD2019P4) x3,21q22.3(RP11-282I20-RP11-1000I21)x3 dn.

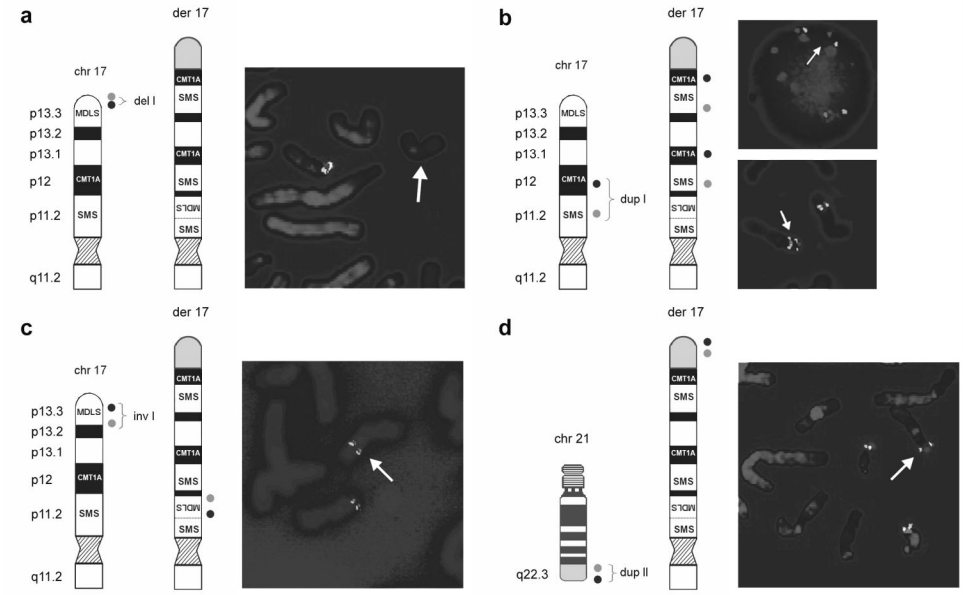


Figure 6.3
Schematic representation of a normal 17p and der(17) (black & white) with translocated chromosome 21 material (blue). The location of the FISH probes are shown on the left side of each figure panel; der(17) is indicated on FISH pictures by a white arrow. (**a**) Terminal deletion of 17pter was validated using BAC clones RP11-1260E13 (red) and CTD-2326F1 (green). (**b**) FISH with *PMP22*-specific PAC RP1-150M12 (red) and *RAI1*-specific BAC RP11-525O11 (green) revealed direct duplication of the CMT1A and SMS regions in 17p12p11.2. (**c**) FISH with PAC RP1-95H6 (red) and BAC GS-202L17 (green) showed inverted insertion of the MDLS region into the SMS region. (**d**) Array CGH also identified a duplication of 21q22.3. Additional FISH analysis using BAC clones RP11-40L10 (green) and RP11-16B19 (red) revealed that the duplicated material of 21q22.3 was translocated onto der(17). Summary of FISH results is provided in **Table 6.1**.

Cytogenetic and molecular analyses of patient 2

Conventional cytogenetic analysis at the 550-band resolution in patient 2 (**Figure 6.1d-e**) showed a normal 46,XX karyotype. Multiplex ligation-dependent probe amplification revealed normal disomic copy numbers of the subtelomeric regions. Subsequent array CGH revealed four interspersed microduplications involving the short arm of chromosome 17, in total comprising 8.8 Mb of genomic sequence (**Table 6.2**). *De novo* occurrence was established by a similar array CGH analysis with DNA obtained from the accompanying parents (**Figure 6.2c**; de Vries et al. 2005). Duplication I in 17p13.2 was determined as ~2.9 Mb in size; duplication II in 17p13.1 was assessed as ~1.4 Mb in size; duplication III in 17p12 was estimated as ~2.9 Mb in size and included the *PMP22* gene known to cause CMT1A disease with dosage-specific overexpression of *PMP22* (Patel et al. 1992). Duplication IV in 17p11.2 was determined as being ~1.5 Mb in size.

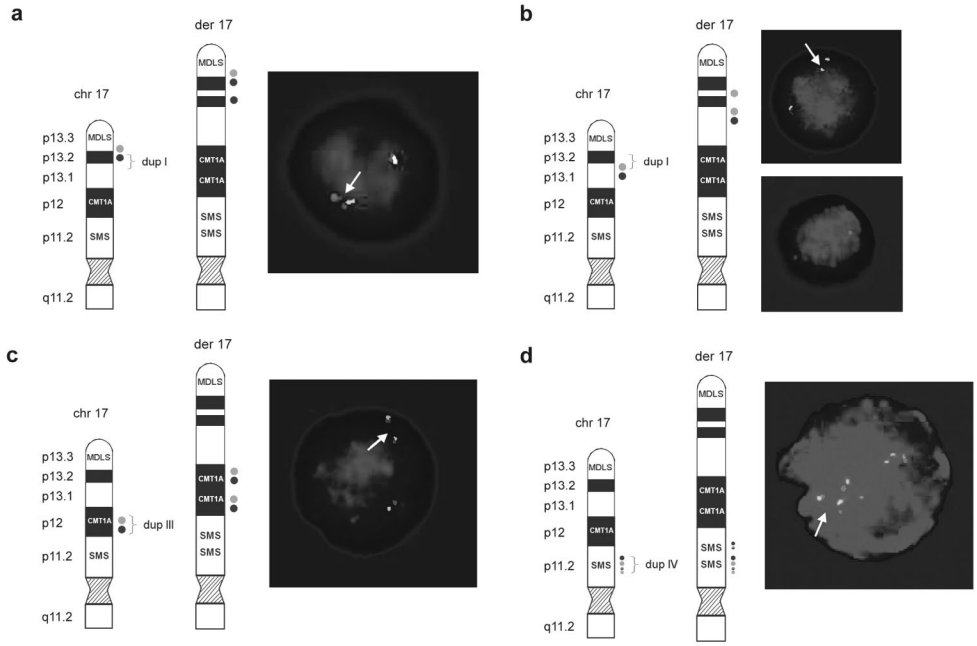


Figure 6.4
Schematic representation of the normal 17p and der(17). The locations of the FISH probes are shown on the left side of each figure panel. (**a**) Duplication I showed relatively simple fluorescence signal pattern with probes RP11-810M2 (green; normal) and RP11-597I9 (red; duplicated) for the distal breakpoint, and (**b**) RP11-222J21 (green; duplicated) and RP11-98D15 (red; normal) for the proximal breakpoint. (**c**) Direct orientation of duplication III was shown using BAC clones RP11-601N13 (green) and RP11-726O12 (red). (**d**) For the distal breakpoint of duplication IV, BAC clones RP11-448D22 (green) and CTD-2145A24 (red) showed duplicated signals on der(17), indicating that both middle SMS-REP and LCR17pB are duplicated as a block. Note four red signals on der(17) representing two normal and two duplicated copies of LCR17pA/B (dupIII) and LCR17pB (dupIV) and four green signals depicting three normal copies of SMS-REPs and the duplicated middle SMS-REP. Summary of FISH results is provided in **Table 6.2**.

Dual color FISH analysis was used to independently confirm the array CGH results (**Figure 6.4; Table 6.2**). In addition, the parental chromosomes were evaluated for predisposing rearrangements within chromosome 17p potentially explaining the complexity of the CCR, however, no such rearrangements were observed. The patients karyotype was finally designated as:

> 46,XX.ish dup(17)(p13.1p13.2)dup(17)(p13.1p13.1)dup(17)(p12p12)dup(17)(p11.2p11.2). arr cgh 17p13.1p13.2(RP11-597I9-RP11-222J21)x3,17p13.1(RP11-63C7-RP11-324C22)x3, 17p12(RP11-333I5-RP11-640N23)x3, 17p11.2(RP11-304M17-CTD-2022C7)x3 dn.

Table 6.2: Summary of the array CGH and FISH results patient 2

Chr. band	Clone[a]		Distance from pter (Mb) [b]	Der (17)	Technique(s)	
17p13.2	RP11-810M2	*G*	4.0	normal	array CGH/ FISH	
17p13.2	RP11-597I9	*R*	4.1	++	array CGH/ FISH	
17p13.2	RP11-106A7		4.3	++	array CGH/ FISH	
17p13.2	RP11-217O7		4.3	++	array CGH	
17p13.1	RP11-816H10		6.7	++	array CGH	dup I
17p13.1	RP11-417F20		6.9	++	array CGH/ FISH	
17p13.1	RP11-558E15		7.1	++	array CGH	
17p13.1	RP11-222J21	*G*	7.2	++	array CGH/ FISH	
17p13.1	RP11-98D15	*R*	7.3	normal	array CGH/ FISH	
17p13.1	RP11-205D17		7.3	normal	array CGH	
17p13.1	RP13-626G5		7.4	normal	array CGH	
17p13.1	RP11-63C7		7.7	++	array CGH	
17p13.1	RP11-441N13		7.8	++	array CGH/ FISH	
17p13.1	RP11-452D1		8.1	++	array CGH/ FISH	
17p13.1	RP11-111I3		9.0	++	array CGH/ FISH	dup II
17p13.1	RP11-85B7		9.1	++	array CGH	
17p13.1	RP11-342E3		9.1	++	array CGH/ FISH	
17p13.1	RP11-324C22		9.2	++	array CGH	
17p13.1	RP11-482E13		9.2	normal	array CGH/ FISH	
17p13.1	RP11-589N11		9.2	normal	array CGH	
17p12	RP11-560N10		12.3	normal	array CGH/ FISH	
17p12	RP11-674I3		12.5	normal	array CGH	
17p12	RP11-746E8		12.5	++	FISH	
17p12	RP11-333I5		12.6	++	array CGH/ FISH	
17p12	RP11-601N13	*G*	13.9	++	FISH	
17p12	RP11-686G16		15.1	++	Array CGH	dup III
17p12	RP11-726O12	*R*	15.2	++	Array CGH/ FISH	
17p12	RP11-385D13		15.4	++	FISH	
17p12	RP11-640N23		15.5	++	Array CGH	
17p12	RP11-640I15		15.6	++	FISH	***LCR17pA***
17p12	CTD-3157E16		15.7	normal	FISH	
17p12	RP11-59N13		15.7	normal	array CGH	

Table 6.2 (continued)

Chr. band	Clone[a]		Distance from pter (Mb) [b]	Der (17)	Technique(s)	
17p11.2	RP11-484D23		18.3	normal	array CGH	
17p11.2	RP11-667E24		18.4	normal	array CGH	
17p11.2	RP1-178F10		18.4	normal	FISH	
17p11.2	RP11-448D22	*G*	18.6	++	FISH	***Middle SMS-REP/ LCR17pB***
17p11.2	RP11-304M17		18.6	++	array CGH	
17p11.2	CTD-2145A24	*R*	18.7	++	FISH	
17p11.2	RP11-137E6		20.0	++	array CGH/ FISH	dup IV
17p11.2	CTD-2022C7		20.0	++	array CGH	
17p11.2	RP11-121A13		20.1	++	FISH	***LCR17pC***
17p11.2	CTD-2313N10		20.3	normal	array CGH	
17p11.2	RP11-185K8		20.4	normal	array CGH	

'++' Duplicated; Extent of duplication is indicated at the right; Breakpoint-associated LCRs are indicated in bold.
[a]: Clones indicated by *R* (for red) or *G* (for green) are shown in **Figure 6.4a-d** in the respective colors
[b]: Distance from pter to the clone midpoints

Overview of the 17p breakpoints in both patients

In total, 17 breakpoints were identified and molecularly characterized at one BAC clone resolution in these two patients, with nine breaks in patient 1 and eight breaks in patient 2. In patient 1, one of the nine breakpoints (inverted insertion of the MDLS region) was mapped within the middle SMS-REP/LCR17pB block after FISH with these LCR flanking clones RP1-178F10 and RP11-28B23. In patient 2, three of the eight breakpoints were associated with LCRs in proximal 17p. The proximal breakpoint of duplication III was located in LCR17pA, between subunits LCR17pA/B and LCR17pA/D using FISH with LCR17pA/B-specific BAC clone RP11-640I15 and LCR17pA/D- and LCR17pA/C- specific BAC clone CTD-3157E16 that are known to cross-hybridize to their homologous copies LCR17pB, LCR17pD and LCR17pC, respectively (Stankiewicz et al. 2004). BAC clone RP11-640I15 showed four fluorescent signals on der(17) (representing two normal and two duplicated copies of LCR17pA/B (dup III) andLCR17pB (dup IV) and CTD-3157E16 three signals on der(17) in analyzed interphase nuclei (depicting normal, not duplicated copies of LCR17pA/D, LCR17pC, and LCR17pD). In duplication IV the distal breakpoints was mapped in the middle SMS-REP, between PAC clone RP1-178F10 (flanks the middle SMS-REP on the telomeric side) and BAC clone RP11-448D22 (middle SMS-REP-specific) (**Figure 6.4d**). The proximal breakpoint of dup IV was mapped in LCR17pC using FISH with its specific BAC clone RP11-121A13. DNA sequence analysis of the genomic regions surrounding the remaining 13 chromosome breakpoints revealed no significant homology or evidence of low-copy repeats. **Figure 6.5** shows an overview of the proximal 17p breakpoints and its association with genomic architectural features observed in both patients. In addition, a selected set of breakpoints previously reported in other patients with 17p aberrations that are associated with LCRs is displayed.

Discussion

The availability of molecular genome profiling techniques such as array CGH have markedly enhanced the resolution of chromosome studies and enabled high resolution genome analysis, thus proving a more accurate method for the identification and delineation of chromosomal rearrangements (Vissers et al. 2003; Shaw-Smith et al. 2004; Cheung et al. 2005; de Vries et al. 2005; Johnston et al. 2007). As a result, precise definitions of CCRs and their true complexity can now be better established. An apparent underestimation of the full complexity of CCRs is well demonstrated in patient 1, in whom a complex karyotype was identified, including an inverted insertion of the MDLS region into the middle SMS-REP/LCR17pB block, two microdeletions (terminal and interstitial in 17p12) and a microduplication involving both SMS and CMT1A chromosome regions. Furthermore, a duplication of 21q22.3qter translocated onto 17p13.3 was identified. The interstitial deletion on 17p12, as well as the duplication of 21q22.3, were not identified prior to array CGH analysis. Thus, by unraveling the complexity of CCRs using array CGH, the added value of this technique to conventional karyotyping was demonstrated. It is, however, noteworthy that current array CGH techniques are incapable of detecting balanced alterations such as inversions, and only provide information regarding genomic gains and losses. FISH can augment the study of CCRs by providing genomic positional and orientational information of imbalances. This is well demonstrated by the apparently balanced inverted insertion of the MDLS region into the middle SMS-REP/LCR17pB block in patient 1, which would have escaped attention if array CGH would have been the only technique employed.

A total of 106 genes in patient 1 and 133 genes in patient 2 were affected by genomic imbalances (gains and losses) because of the CCRs. For patient 1, the number of genes affected is less than expected based on the average number of 15 genes/Mb on chromosome 17 and 6 genes/Mb on chromosome 21. It thus seems that the CCR in patient 1 affected relatively gene poor regions. For patient 2, the total number of genes affected was in concordance with the expected number of genes affected by the CCR. Given the large number of genes affected in each patient, it is difficult to correlate any specific gene with the observed phenotypes. However, a few genes affected by the CCRs are well established dosage-sensitive genes, including *RAI1* in patient 1 and *PMP22* in patient 2, which are known to cause disease when duplicated. Recently, duplication of *RAI1* has been shown to cause a physical and behavioral phenotype termed Potocki-Lupski syndrome (Potocki et al. 2007). However, due to the young age of patient 1, as well as the imbalances of other genomic regions, assessing phenotype-genotype correlations is not possible. One of the genes affected by the CCR in patient 2 is the dosage-sensitive *PMP22* gene. Duplication of this gene causes CMT1A, a common inherited neuropathy characterized by myelin degeneration (Patel et al. 1992). Indeed, the first clinical signs of CMT1A were present in this patient.

Genome profiling techniques provide a readily accessible platform for the delineation of complex rearrangements at the breakpoint level, thus allowing the study of the molecular mechanisms underlying such rearrangements. For recurrent as well as nonrecurrent alterations, it has been shown that genomic architectural features such as LCRs and smaller repetitive elements including *Alu* sequences are capable of mediating and/or stimulating the occurrence of deletions and duplications (Pentao et al. 1992; Chen et al. 1997; Shaw & Lupski 2005). LCRs in proximal 17p comprise more than 23% of the

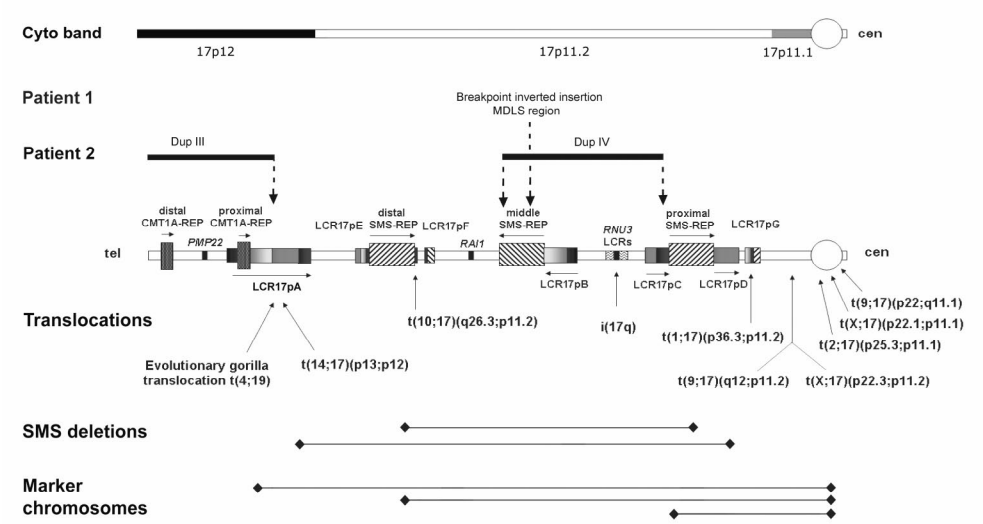


Figure 6.5
Top: Cytogenetic chromosome bands 17p12, and sub-bands 17p11.2 and 17p11.1 are shown. Breakpoints of the chromosome rearrangements in patients 1 and 2 associated with architectural features in proximal 17p are shown by vertical dotted blue arrows. In patient 1, one breakpoint is located within the middle SMS-REP/LCR17pB block, whereas in patient 2 breakpoints are located within LCR17pA (proximal Dup III), middle SMS-REP (distal Dup IV) and LCR17pC (proximal Dup IV). Duplication III and IV in patient 2 are indicated by horizontal blue lines. **Bottom**: Previously identified rearrangements associated with LCRs in 17p. Breakpoints of translocations and isochromosome 17q are indicated by vertical black arrows whereas common ~4 Mb and uncommon ~5 Mb SMS deletions and marker chromosomes are indicated by black horizontal lines. The LCR17p structures are depicted in colors to better represent their positional orientation with respect to each other; the shaded rectangles and horizontal black arrows represent the orientation of the LCRs.

genomic sequence, which is approximately fourfold higher than might be expected based on virtual analysis of the entire human genome (Stankiewicz et al. 2003). It might be anticipated that the breakpoints of the uncommon nonrecurrent chromosomal aberrations as observed in this study occurred by chance alone within one of these LCRs. However, of five breakpoints that occurred within the proximal part of 17p, four were located within a LCR structure. The remaining 10 breakpoints were located in the LCR-poor distal fragment of 17p. Thus, it appears that the location of the breakpoints is associated with local genome architecture, i.e, LCRs. Based on the abundance of genomic architectural features in proximal 17p, it remains to be determined whether the two CCRs described in this study are representative for other CCRs. However, a comparable genomic complexity has been observed in other genomic regions. For instance, LCRs in 22q11.2 have been shown to be responsible for recurrent and nonreccurrent chromosome deletions, duplications, and translocations (Shaikh et al. 2001; Spiteri et al. 2003). Thus, it can be expected that additional complex rearrangements with genomic architecture playing a mediating role, will be identified.

The molecular characterization of the complex rearrangements enabled the identification of nine and eight breakpoints in the CCRs in patient 1 and 2, respectively. Of all these 17 breakpoints, four involved known LCRs being the middle SMS-REP/LCR17pB block in patient 1 and the middle SMS-REP, LCR17pA and LCR17pC in patient 2. The remaining breakpoints did not reveal any significant similarity and/or any evidence for the presence of LCRs. The proximal breakpoint of duplication III in patient 2 was located within LCR17pA. This genomic region has previously been found to be involved in several other constitutional and evolutionary rearrangements, including t(4;19) in *Gorilla Gorilla* (Stankiewicz et al. 2001a; Stankiewicz et al. 2001b; Stankiewicz et al.2003; Stankiewicz et al. 2004; Shaw et al. 2004a; Lupski & Stankiewicz 2005; Yatsenko et al. 2005; Ou et al. 2006). The distal breakpoint of this duplication, however, was not located within a genomic segment sharing sequence homology with LCR17pA. Breakpoints of duplication IV were mapped within non-homologous LCR copies, middle SMS-REP and LCR17pC. The middle SMS-REP/LCR17pB was also insertion target of the MDLS region in patient 1. These observations suggest that other recombination mechanisms, such as nonhomologues end-joining (NHEJ), may have facilitated these CCRs. NHEJ may play a prominent role in nonrecurrent rearrangements as has been shown for *PLP1* deletions and duplications (Inoue et al. 2002; Lee et al. 2006). It remains to be determined to what extent NHEJ is a mechanism for genomic rearrangements, since other potential replication-based mechanisms may occur with nonrecurrent rearrangements (Lee et al. 2006).

The detailed molecular information gathered on the breakpoints does, however, allow for speculation on the mechanism underlying CCR formation in our patients. We propose models based upon the principle of parsimony and the minimal amount of breaks required for formation of the CCRs. Although predisposing inversion(s) and translocations rendering susceptibility to rearrangements in the offspring similar to what is known for several genomic disorders have been excluded in the parental chromosomes (Axton 2006; Lupski 2006), alternative mechanisms of CCR formation cannot be excluded fully. The complex rearrangement observed in patient 1 includes nine breakpoints, with one insertion, two microdeletions, one microduplication and one translocation. We propose that the breaks in der(17), 17p13.3, 17p12 and 17p11.2 as well as the break in chromosome 21q22.3 arose simultaneously. This resulted in the insertion of the MDLS region into the middle SMS-REP/LCR17pB block, loss of the subtelomeric region 17p13.3 and part of 17p12, and duplication of the CMT1A and SMS regions. Additionally, the duplicated 21q22.3 region was translocated to the 17p subtelomeric region (**Figure 6a**), most likely by a telomere healing mechanism. The complex rearrangement in patient 2 includes eight breaks. For the most parsimonious mechanism, a single breakpoint is involved in both, the formation of, as well as the insertion of, the duplicated segment. Thus, for each duplication observed in patient 2, one of the breakpoints is involved in the formation of the duplication and provides at the same time the site for inserting the duplicated segment. We suggest that also in this case all breaks in chromosome 17 occurred concurrently, resulting in four interspersed directly orientated microduplications (**Figure 6b**).

In conclusion, the human genome contains many genomic architectural features such as LCRs, of which several have been identified at the breakpoints of (recurrent) chromosome rearrangements. As such, these chromosome rearrangements do not appear to represent random events but, instead, result from underlying genomic architectural features. Here, we present two patients with unique *de novo* complex chromosome rearrangements, of

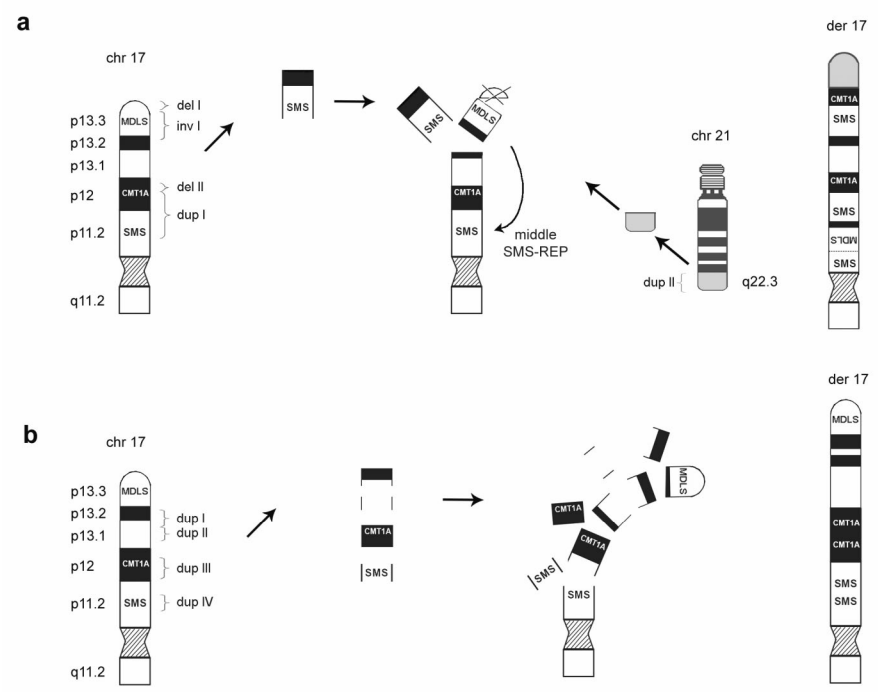


Figure 6.6
Schematic representation of the proposed mechanism for CCR formation in patient 1 and patient 2. (**a**) The complex rearrangement observed in patient 1 includes ten breakpoints, with one insertion, two microdeletions and two microduplications. We propose that the breaks in chromosome 17, p13.3, p12 and p11.2 as well as the break in chromosome 21q22.3 arose simultaneously. This resulted in the insertion of the MDLS region into the middle SMS-REP/LCR17pB block, loss of telomeric 17p and part of 17p12 and duplication of the CMT1A region. Additionally, the duplicated 21q22.3 fragment was translocated to the 17p subtelomeric region. (**b**) The complex rearrangement in patient 2 included eight breaks. We suggest that breaks in chromosome 17, p13.2, p13.1, p12 and p11.2 occurred at the same time, resulting in four interspersed directly orientated microduplications. Note: arrows do not represent a chronological order of events.

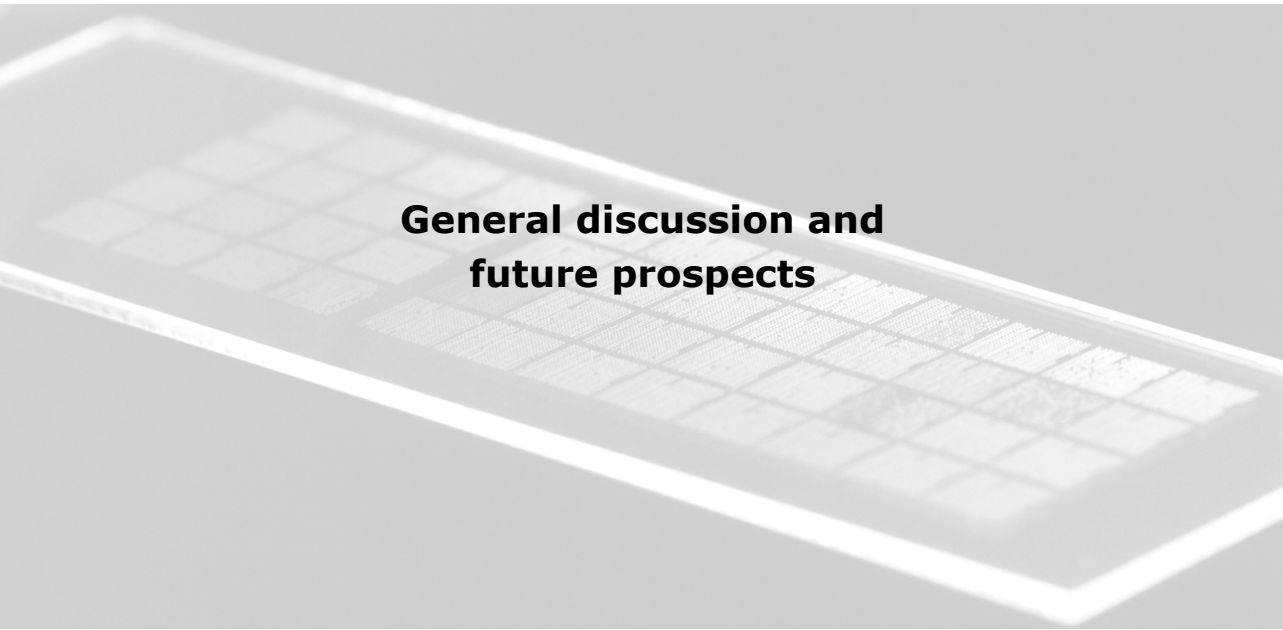
which several breakpoints are located within LCRs. These results provide evidence for the first time that in addition to previously reported chromosome deletions, duplications, inversions, translocations and marker chromosomes, genomic architectural features such as LCRs are mechanistically important for the origin of some complex chromosome rearrangements.

Acknowledgements

We thank the patients and their parents for participation in this research, and Margie Withers, Irene Janssen, Saskia van der Velde-Visser and Christel Beumer for excellent technical assistance. This work was supported by the Netherlands Genomics Initiative (050-72-407) (LELMV), the Netherlands Organisation for Health Research and

Development (907-00-058; BBAdV and 912-04-047; JAV), the National Institute of Child Health and Development (PO1 HD39420) (JRL) and the Baylor College of Medicine Mental Retardation Research Center (HD24064).

Chapter 7



General discussion and future prospects

Until recently, it was impossible to analyze the entire human genome for the presence of submicroscopic copy number variations. The development and subsequent implementation of various microarray-based genome profiling technologies such as array CGH, however, has enabled such whole-genome analyses at a resolution that vastly exceeds that of conventional karyotyping. In this chapter, several aspects of this revolutionizing technology will be discussed, including its implications for current and future clinical applications and genome research.

7.1 The continuously increasing resolution of genome profiling technologies

Genome profiling technologies, such as array CGH, have dramatically changed the nature of human genome analysis by combining the targeted high-resolution approach of the FISH technology and the whole-genome approach of the karyotyping technology. Many laboratories have started their array CGH studies using low-density custom made arrays encompassing probes specifically targeted at selected genomic regions. Examples of these are arrays targeting all subtelomeric regions (Veltman et al. 2002; Harada et al. 2004), arrays targeting regions known to be involved in 'microdeletion/microduplication syndromes' (Yu et al. 2003; Klein et al. 2004; Locke et al. 2004; Shaw et al. 2004b; van Buggenhout et al. 2004; Wang et al. 2004), or arrays targeting other chromosomal regions of interest (Buckley et al. 2002; Veltman et al. 2003a; Ekong et al. 2004; Solomon et al. 2004; Veltman et al. 2004; Koolen et al. 2005). The ultimate power of the microarray technology for copy number assessment lies, however, in the ability to screen the entire human genome, without *a priori* knowledge of the genomic region involved. To this end, high-density arrays were constructed covering the entire human genome, initially with a 0.75 to 1.4 Mb resolution using ~3,000 BAC clones and, more recently, with a resolution of 50-100 kb using ~32,000 BAC clones (Snijders et al. 2001; Vissers et al. 2003; Shaw-Smith et al. 2004; Ishkanian et al. 2004; de Vries et al. 2005). Although these arrays have found a wide application in genetic diagnostics and research (see paragraphs 7.2-7.4), the resolution of these microarrays does not allow analysis at the level of individual genes, let alone individual exons. Thus, efforts were made to further increase the resolution of existing genome profiling technologies.

The subsequent use of alternative array elements such as oligonucleotides targeting random sequences or SNPs have indeed taken this resolution to a next level (Barrett et al. 2004; Larrabee et al. 2004; Rauch et al. 2004; Schaeffer et al. 2004; Schoumans et al. 2004, Herr et al. 2005; Slater et al. 2005; Friedman et al. 2006; Peiffer et al. 2006). Currently, the highest density oligonucleotide-based microarrays encompass over 500,000 probes, resulting in an average genome-wide spacing of one oligonucleotide every 6 kb. In the near future, this resolution is expected to increase even further with the commercial availability of microarrays encompassing a million or more probes. Recently developed custom made arrays have already enabled a reliable copy number assessment at a resolution <1 kb for targeted genomic regions. These include arrays targeting exon sequences from a limited set of known disease genes (Dhami et al. 2005) and arrays targeting known microdeletion syndromes (Sharp et al. 2006; Johnston et al. 2007).

In addition to its high resolution, commercially available microarrays have several advantages over in-house produced microarrays such as (i) the minimization of costs due to the large-scale production, (ii) the wide-spread availability of microarrays to research

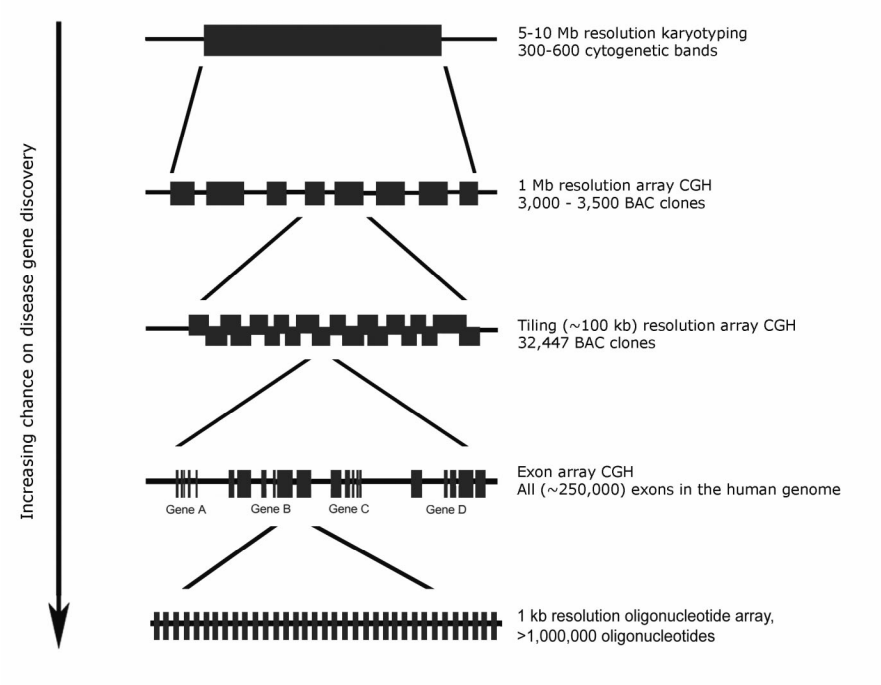


Figure 7.1
Impact of the increasing resolution of molecular karyotyping on disease gene identification.

and diagnostic laboratories (iii) the availability of large (combined) data sets from control and test samples and (iv) the possibility to directly compare data sets from different laboratories (Veltman 2006). An additional advantage of oligonucleotide-based SNP arrays is the ability to simultaneously measure variations in signal intensity and allelic composition, thus allowing the detection of both copy number variation and copy-neutral loss-of-heterozygosity (LOH) events. The detection of copy-neutral LOH, i.e., the detection of LOH without copy number reduction, has already successfully been applied to the genomic analysis of chronic lymphocytic leukaemia (Pfeiffer et al. 2007) and medulloblastoma (Langdon et al. 2006). It is anticipated that such analyses will increasingly be applied to the identification of epigenetically regulated and/or imprinted genes that may be linked to disease.

Apart from increasing resolution for the detection of copy number variation, technological developments are ongoing to interrogate the human genome at the highest level of resolution possible, i.e., at the single base pair level. Currently, the key limitations for individual whole-genome re-sequencing are the high costs involved and relative low throughput. For example, if performed today, an individual human genome sequence would cost approximately €10 million, and it would take 30 standard sequencing instruments 1 year to finish the sequence (Bentley 2006). Novel whole-genome sequencing platforms such as those from 454 Life Sciences and Solexa, will dramatically increase throughput and lower the costs, thus bringing the concept of personalized

genome sequencing within reach. Such personalized genome sequence will not only facilitate the identification of single gene mutations, but may also allow for copy number detection and the identification of (submicroscopic) balanced rearrangements.

In conclusion, microarrays with a resolution of 1 kb, or even beyond, will soon become available on a genome-wide scale. In combination with upcoming whole-genome sequencing approaches, all chromosomal rearrangements, including submicroscopic balanced rearrangements, will be detectable in the near future. It is envisioned that this increase in resolution will lead to a concomitant increase in disease gene discovery (**Figure 7.1**).

7.2 Diagnostic molecular karyotyping

In chapter 2 we have shown that array CGH-based molecular karyotyping, using genome-wide 1 Mb BAC arrays, allows the detection of causative microdeletions and/or duplications in 10% of individuals with mental retardation with or without additional congenital malformations. Additional studies have provided insight into the quality and reproducibility of the procedure, the need for validation of the microarray data by independent technologies such as FISH or MLPA, as well as the way to translate these data into clinical practice (Shaw-Smith et al. 2004; Schoumans et al. 2005; Tyson et al. 2005). The clinical usefulness of molecular karyotyping was further substantiated in larger, less selected, cohorts of individuals with mental retardation using 1 Mb resolution BAC arrays (Menten et al. 2006; Rosenberg et al. 2006), tiling-resolution BAC arrays (de Vries et al. 2005), or 100k SNP arrays (Friedman et al. 2006). When taking these studies together, two main conclusions can be drawn (i) in addition to submicroscopic subtelomeric chromosome imbalances, submicroscopic interstitial chromosome imbalances are responsible for a considerable proportion of cases with mental retardation varying between 5 and 20% depending on the clinical pre-selection of the individuals, and (ii) these interstitial submicroscopic chromosome imbalances occur all over the genome without preference for specific loci (reviewed by Knight & Regan 2006; Veltman 2006).

Remarkably, we found that the diagnostic yield did not increase when using higher resolution microarrays, i.e., ~1 Mb versus ~100 kb BAC arrays, respectively (Chapter 2; de Vries et al. 2005). It must, however, be noted that the diagnostic yield of this approach heavily depends on the number and quality of previous cytogenetic studies performed as well as on the clinical selection of the patients prior to the array studies. Indeed, the first pilot studies were performed on small series of patients that were stringently selected, leading to an over-estimation of the overall diagnostic yield. The current, and most reliable, estimated diagnostic yield is based on the analysis of 500 karyotypically normal patients with mental retardation in which subtelomeric abnormalities had been excluded, but without any further clinical pre-selection, using a tiling resolution BAC array. This study revealed 48 *de novo* disease-related interstitial submicroscopic copy number imbalances (9.6%) (Koolen et al. *manuscript in preparation*). Approximately half of these imbalances were below the level of <1 Mb, thus underscoring the importance of using arrays beyond the resolution of ~1 Mb. As stated above, experimental and technical aspects will ultimately no longer be the limiting factors in the detection of (submicroscopic) deletions and duplications. So far, the resolution to detect chromosome rearrangements

has increased 100-fold, that is, from a 10 Mb resolution using conventional karyotyping to a 100 kb resolution using molecular karyotyping by tiling resolution BAC arrays. However, the ultimate resolution at a single base pair level will be another four orders of magnitude larger than the current resolution. The above estimated diagnostic yield of ~10% is therefore undoubtedly an underestimation for interstitial submicroscopic rearrangements causative for mental retardation.

The increase in resolution of genome analysis technologies has also led to the identification of frequently occurring large scale genomic imbalances, denoted as copy number variations (CNVs), which may be inherited from unaffected parents. In our first study, using a ~1 Mb microarray, we identified three submicroscopic rearrangements representing such CNVs in a series of 20 patients (Chapter 2). More recently, de Vries et al. (2005) identified 258 CNVs inherited through one of the apparently healthy parents in 99 of 100 patients with mental retardation using a tiling resolution BAC array. In order to discriminate between causative alterations and benign CNVs, the validation for *de novo* occurrence has become inevitable. Together, these results indicate that the ultimate challenge of using high-resolution arrays for molecular karyotyping lies in the clinical interpretation of the data obtained, as the majority of copy number imbalances detected will represent CNVs inherited from one of the phenotypically normal parents (see also paragraph 7.6).

When a new cytogenetic diagnostic technique replaces an old(er) technique, the new technique needs to have added value. Approximately 5% of mental retardation is caused by chromosomal rearrangements detectable by conventional karyotyping (de Vries et al. 2005). Molecular karyotyping has already shown added value by increasing this percentage to ~20%, including the detection of 2-3% microscopically visible unbalanced rearrangements, 6% submicroscopic subtelomere rearrangements, and 10% interstitial submicroscopic rearrangements. Driven by the commercial availability of genome-wide high-density microarrays, combined with automation of the procedure, it is expected that molecular karyotyping will soon find widespread diagnostic application.

Currently, molecular karyotyping is restricted to postnatal analysis whereas conventional karyotyping is also used in a prenatal setting. There may, however, be two important advantages of using molecular karyotyping, (i) the more reliable detection of subtle rearrangements that may be difficult to detect on prenatal samples, i.e., small telomeric abnormalities, and (ii) the detection of abnormalities that are not detectable even by an optimal karyotype, i.e., cryptic telomeric deletions and duplications or microdeletions and duplications leading to known syndromes (Sahoo et al. 2006). It is, however, questionable whether genome-wide microarrays should be used in a prenatal setting as the clinical interpretation of the vast majority of the (submicroscopic) *de novo* copy number variations as well as inherited copy number variations remains elusive (see also paragraph 7.6). However, targeted arrays to exclude the presence of (cryptic) telomere rearrangements and known microdeletion syndromes could serve as an alternative. Although the incidence of the individual disorders tested on such targeted arrays are rare, their combined live-birth incidence is significant and could be as high as 1:500. Moreover, such targeted arrays have already been successfully applied in prenatal diagnosis (Sahoo et al. 2006).

In conclusion, submicroscopic subtelomeric and/or interstitial chromosome rearrangements together explain up to 20% of the cases of mental retardation, whereas

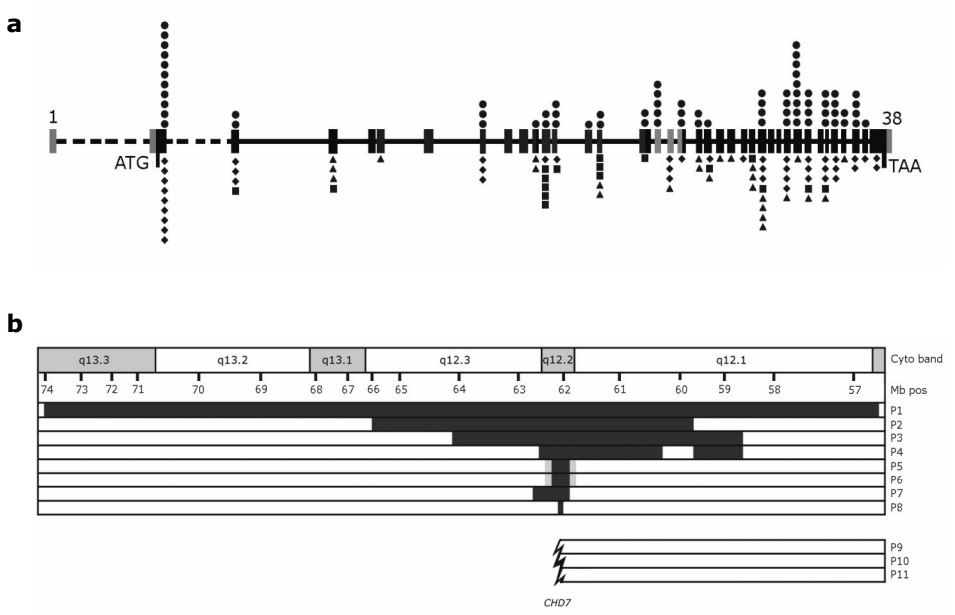


Figure 7.2
Schematic overview of *CHD7* mutations and chromosomal rearrangements detected in patients with CHARGE syndrome. (**a**) Nonsense mutations (circles) are depicted above the schematic representation of the *CHD7* gene, whereas missense (square), frameshift (diamond) and splice site (triangle) mutations are depicted below. Nonsense, frameshift and splice site mutations are scattered throughout the gene, whereas missense mutations are concentrated in the SNF2 domain. (**b**) Deletions ranging from microscopic (P1) to submicroscopic (P8) are indicated for patients 1-8 (P1-P8). Translocation partners disrupting the *CHD7* gene included chromosomes 2, 6, and 13 (P9-P11). Chromosomal bands are indicated by grey and white blocks. Deleted chromosomal regions are represented by black bars, whereas white bars indicate normal chromosome regions.

only 5% of the cases is explained by microscopically visible chromosomal rearrangements. Thus, it is anticipated that this technology will soon replace routine chromosome analysis in the postnatal diagnosis of patients with mental retardation. In addition, it may be expected based on its successful application to mental retardation that molecular karyotyping will soon be used as a screening tool for microdeletions and microduplications in other diseases as well, including common complex and/or multifactorial diseases such as asthma, epilepsy, autism, rheumatoid arthritis and schizophrenia.

7.3 Resolving the genetic cause of known syndromes

In addition to screening individuals with mental retardation, high-resolution genome profiling technologies may also facilitate the identification of disease genes underlying known syndromes for which the genetic cause has remained elusive. The specific phenotype observed in patients with such syndromes allows for a stringent pre-selection of patients whose DNA can subsequently be interrogated using such high-resolution genome

profiling techniques. In Chapter 3 we describe the localization and identification of the gene underlying CHARGE syndrome, through the detection and characterization of microdeletions by array CGH. Sequence analysis of nine genes located within the minimal region of deletion overlap revealed causative mutations in *CHD7*, a novel member of the chromodomain helicase DNA-binding gene family, in the majority of individuals with CHARGE syndrome without deletions. From these results, we concluded that CHARGE syndrome is caused by haploinsufficiency of the *CHD7* gene, either by microdeletions encompassing the *CHD7* gene, or by mutations within this gene.

After our initial identification of *CHD*7 as the causative gene for CHARGE syndrome, four follow-up studies have reported on the frequency and spectrum of *CHD7* mutations in over 200 individuals with CHARGE syndrome (Aramaki et al. 2006; Jongmans et al. 2006; Lalani et al. 2006; Sanlaville et al. 2006). In this series, an overall detection rate of *CHD7* mutations of ~65% was achieved. The mutations were scattered throughout the coding exons and conserved splice sites of *CHD7* (**Figure 7.2a**). The majority of mutations represented nonsense or frame shift mutations (74%), predicted to result in premature termination of the protein, most likely leading to haploinsufficiency. However, we assume that the mutation detection rate of 65% is an underestimation, as three of four studies were not designed to identify exonic deletions (Jongmans et al. 2006; Lalani et al. 2006; Sanlaville et al. 2006) and, in addition, all four studies included patients not strictly fulfilling the CHARGE criteria (Blake et al. 1998). In one study, 38 of such 'CHARGE-like' patients were included for *CHD7* mutation analysis (Jongmans et al. 2006). In none of these patients mutations could be identified. Exclusion of such 'CHARGE-like' patients from this study would increase the mutation rate to 96% for patients that strictly fulfil the CHARGE criteria.

Although *CHD7* point mutations are the main cause of CHARGE syndrome, several additional microdeletions encompassing the *CHD7* gene have been detected (**Figure 7.2b**). These deletions range from a submicroscopic intragenic deletion of exons 8-12 (Aramaki et al. 2006) to a large cytogenetically visible ~18 Mb deletion encompassing the region 8q11.2-q13 (Arrington et al. 2005). In addition to *CHD7* deletions, also translocations disrupting the *CHD7* gene have been identified, including a t(8;13)(q11.2;q22) in a monozygotic twin pair (Johnson et al. 2006). The frequency of (sub)microscopic chromosome rearrangements (deletions and translocations) of *CHD7* in individuals with CHARGE syndrome is at present ~5% (Vissers et al. *unpublished results*).

As yet, the exact biological function of CHD7 remains to be established. However, like its family members, CHD7 is thought to play an important role in chromatin remodeling during early development, allowing a level of epigenetic control over target genes expressed in mesenchymal cells derived from the cephalic neural crest. Embryonic expression profiling studies of mouse and chicken *Chd7* showed expression in exactly those tissues that are affected in human CHARGE syndrome patients (Bosman et al. 2005; Aramaki et al. 2007). Moreover, mRNA *in situ* hybridization analysis of the *CHD7* gene during early human development emphasized a role of CHD7 in the development of the central nervous system, inner ear, and neural crest of pharyngeal arches, thus, underscoring a good correlation between the *CHD7* expression pattern and the various developmental anomalies observed in CHARGE syndrome (Sanlaville et al. 2006).

Table 7.1: Recently defined genomic syndromes, their chromosomal localization and genes involved

Syndrome	Chromosome	Gene(s) involved	Size	Detectable by tiling resolution BAC array[a]	Detectable by 500K SNP array[a]	Reference
1q21	del(1)(q21.1)	tbd	718 kb	+++	++++	de Vries et al. 2005; This chapter
Cystinuria with mitochondrial disease	del(2)(p16)	*SLC3A1; PPM1B; KIAA0436*	179 kb	++	+++	Parvari et al. 2001
2q22.3-q23	del(2)(q22.3-q23) dup(2)(q23.1q23.1)	tbd	diverse	+++	++++	Vissers et al. 2003; de Vries et al. 2005; This chapter
3q29 deletion syndrome	del(3)(q29)	tbd	1500 kb	+++	++++	Willatt et al. 2005
Adrenal hyperplasia with hypermobility	del(6)(p21)	*TNBX; CYP21A*	33 kb	-	++	Koppens et al. 2000
9q22.3 deletion syndrome	del(9)(q22.3)	tbd	6500 kb	+++	++++	Redon et al. 2006b
Otofacialcervical syndrome	del(8)(q13.3)	EYA1	316 kb	++	+++	Rickard et al. 2001
Potocki-Shaffer	del(11)(p11.2)	*EXT2; ALX4*	2100 kb	+++	++++	Potocki et al. 1996
Infantile hyperinsulinism, enteropathy and deafness	del(11)(p15p14)	*USH1C; ABCC8; KCNJ11*	122 kb	+	++	Bitner-Glindzicz et al. 2000
12q14 microdeletion syndrome	del(12)(q14)	*LEMD3; HMGA2; GRIP1*	3440 kb	+++	++++	Menten et al. 2007
15q24	del(15)(q24.1q24.2)	tbd	1650 kb	+++	++++	de vries et al. 2005; Sharp et al. 2007
Tuberous sclerosis, polycystic kidney disease	del(16)(p13)	TSC2; PKD1	87 kb	+/-	++	Brook-Carter et al. 1994
Potocki-Lupski syndrome	dup(17)(p11.2p11.2)	*RAI1*	3700 kb	+++	++++	Potocki et al. 2007
17q21.31 microdeletion syndrome	del(17)(q21.31)	tbd	600 kb	++	++++	Chapter 4
Alport-leiomyomatosis	del(X)(q22.3)	*COL4A5; COL4A6*	133 kb	+	++	Zhou et al. 1993

[a]:detection indicated by '+' and '-' signs for detectable and undetectable, based on genome-wide tiling resolution BAC array or 500K SNP array, respectively
tbd: to be determined.

Placed in a broader perspective, molecular karyotyping may be best suited for resolving the genetic cause of known syndromes that involve haploinsufficiency as the disease causing mechanism. Whether this latter is the case, may be impossible to predict from the phenotype alone. For example, much effort went into a strategy that aimed at the identification of the gene that causes Noonan syndrome by detecting deletions in individual patients with a Noonan-like phenotype (Caballin et al. 1891; Onufer et al. 1987; Robin et al. 1995; Ion et al. 2000). This strategy, however, failed because all causative mutations of the *PTPN11* gene and *KRAS* gene turned out to be missense mutations (Tartaglia et al. 2001; Schubbert et al. 2006). A further constraint on the use of deletion and duplication screening for disease gene identification is related to the local genome composition. The frequency of patients with large rearrangements appears to depend on the sequence characteristics of the region involved, which may contain repeats such as LCRs that predispose to deletion (Stankiewicz & Lupksi 2002; paragraph 7.5). Another relevant consideration is the presence of further genes in the region that are subject to gene dosage effects. Significant mental retardation, for instance, predicts the presence of microdeletions for a number of single gene conditions, such as Rieger syndrome (de Baere et al. 2003), Saethre-Chotzen syndrome (Johnson et al. 1998) and Aniridia Type II (Ton et al. 1991; Crolla & van Heyningen 2002). In addition, combinations of clinical features may occur through contiguous gene syndromes, which continue to be defined (paragraph 7.4, **Table 7.1**). In the case of CHARGE syndrome, the index patient with the ~5 Mb deletion presented with relatively severe mental retardation (Chapter 3). Thus, selection of individual cases presenting with mental retardation and additional clinical features will increase the chance of disease gene discovery.

The relatively low frequency of microdeletions in CHARGE syndrome indicates that it is difficult to predict *a priori* how many patients need to be tested before submicroscopic rearrangements are encountered which, subsequently, may lead to disease gene discovery. This is nicely illustrated by the discovery of *NMYC* as the causative gene for Feingold syndrome. A 4 Mb microdeletion was identified in one out of five families with the Feingold syndrome that were studied for linkage on chromosome 2 (Celli et al. 2000). Haploinsufficient point mutations of the *NMYC* gene were subsequently identified in several additional families, but also a second 1.2 Mb microdeletion, thus yielding a provisional estimate of 10% occurrence for microdeletions in this syndrome (van Bokhoven et al. 2005).

In conclusion, microdeletions and/or microduplications may comprise up to 15% of all mutations underlying monogenic diseases. Thus, molecular karyotyping can be used as a powerful disease gene identification strategy, especially when straightforward linkage mapping is impractical or impossible due to reproductive lethality. This strategy may become even more succesfull when combined with whole-genome sequencing.

7.4 Unravelling the genetic cause of new syndromes

The possibility to screen patients with mental retardation in a high throughput manner at a genome-wide scale using array CGH has substantially increased the chance to identify novel 'microdeletion/microduplication syndromes'. In general, the identification of novel syndromes is based on an accurate phenotype-genotype correlation. From an historical perspective, this correlation relied on a detailed and accurate phenotypic description of the

patients after which overlapping chromosomal rearrangements were uncovered. Nowadays, obtaining a genotype has become much easier and, in addition, has never been more accurate. As a result, the identification of novel syndromes may start with the identification of overlapping genotypes, i.e., the identification of overlapping microdeletions and/or duplications in heterogeneous patient groups. Such overlapping microdeletions and/or duplications are most likely identified in large patient cohorts screened in a high-throughput diagnostic setting using high-resolution genome profiling technologies (see above 7.2). This approach has been proven successful as reported in Chapter 4. Screening of 1,200 individuals with mental retardation in a diagnostic setting using a tiling resolution microarray and MLPA analysis revealed three patients with similar-sized overlapping microdeletions on 17q21.31. Subsequent detailed clinical examination of these patients revealed a clear phenotypic overlap, thus defining a new syndrome. The overlapping deletions encompass six known genes, including *CRHR1*, and *MAPT.* Both latter genes are highly expressed in brain and have been implicated in several neurodegenerative and behavioral phenotypes. These characteristics turn these genes into excellent candidates for dosage-sensitive genes underlying the 17q21.31 microdeletion syndrome (see also paragraph 7.5).

Interestingly, the 17q21.31 microdeletion syndrome is associated with the presence of an inversion-bearing haplotype (H2 allele), in which low-copy repeats are re-organized in such way that they predispose the chromosome to microdeletions in meiotic recombination. The 17q21.31 inversion haplotype is carried at a frequency of ~20% in populations of European ancestry, with evidence that it is under positive selection. In other populations a lower frequency for this H2 allele is observed (Stefansson et al. 2005). Rearrangements mediated by this structurally different haplotype could account for up to 1% of the total population of mental retardation, but with the inversion haplotype carried at different frequencies in different ethnic populations, the microdeletion frequency is likely to differ as well (Axton 2006). Predisposing parental inversion are a common finding in other microdeletion syndromes, such as Angelman syndrome, Williams Beuren syndrome and Sotos syndrome (Osborne et al. 2001; Gimelli et al. 2003; Visser et al. 2005). The breakpoints of the inversion polymorphism are identical to the breakpoints of the common microdeletion, and are thus flanked by low-copy repeat structures (Gimelli et al. 2003). As such, the predisposing inversion can be considered a genomic polymorphism that facilitates the occurrence of the disease-associated rearrangement. Due to the emerging availability of high-resolution genome profiling technologies, it is expected that such predisposing submicroscopic inversions will be identified at an increasing rate.

This newly identified 17q21.31 microdeletion syndrome is not the only example of a novel syndrome identified by high-resolution, genome-wide profiling technologies. Recurrent overlapping submicroscopic rearrangements have also been identified at chromosome 1q21.1, 2q22.3-q23, 5q35.1, and 15q24, for which assessment of the phenotypic overlap in the patients is currently ongoing (Sharp et al. 2007; **Table 7.1**). Together, these findings emphasize the need to search for additional *de novo* genomic rearrangements as a cause of congenital abnormalities (Lupski 2006) and indeed confirm the use of genome-profiling technologies as a tool for the elucidation of such novel syndromes. In order to facilitate genotype-phenotype correlations and the identification of additional syndromes, databases have been designed, including (i) ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations), a database collecting and

providing cytogenetic and clinical information on rare chromosomal disorders including microdeletions and microduplications (Feenstra et al. 2006; www.ecaruca.net/), and (ii) DECIPHER (DatabasE of Chromosome Imbalance and Phenotype in Humans Using Ensembl Resources), a database fully dedicated to submicroscopic chromosome rearrangements (www.sanger.ac.uk/PostGenomics/decipher/).

In conclusion, genotype-phenotype correlations are facilitated by the use of genome profiling technologies and can lead to the assignment of specific phenotypic traits to individual genes. Moreover, these detailed genotype-phenotype correlations can lead to the identification of novel syndromes.

7.5 Single gene disorder, contiguous gene syndrome or genomic disorder?

Chromosome rearrangements may occur via two main molecular mechanisms, (i) nonallelic homologous recombination (NAHR) and (ii) non-homologous recombination, including non-homologous DNA end joining (NHEJ). NAHR appears to be the predominant mechanism underlying recurrent rearrangements that are caused by misalignment of LCRs, sharing long regions of high homology. For non-recurrent rearrangements, in which such long stretches of homology are not present, joining of the DNA ends appears to occur predominantly by NHEJ (Shaw & Lupski 2004; Lupski & Stankiewicz 2005; Shaw & Lupski 2005; Lieber et al. 2006; Lupski & Stankiewicz 2006). In order to classify a newly identified syndrome as a single gene disorder, a contiguous gene syndrome or a genomic disorder, the first step is to determine which of the above mechanisms has mediated the genomic rearrangement. A first clue for this may come from the genomic size and the exact breakpoint locations of the identified rearrangements. If the genomic size of the rearrangement in different patients is largely identical, and the breakpoints cluster together, it is likely that genomic architectural features such as LCRs or other repeat sequences (*Alu* sequences, LINE and SINE elements) have predisposed to genome rearrangements observed (Shaw & Lupski 2004). Next, detailed *in silico* analyses using bioinformatic tools may be used to identify such genomic architectural features located at the breakpoints. These genomic architectural features, if present, should be mapped according their respective orientation since directly orientated repeats may mediate the formation of microdeletions and microduplication through NAHR. In case directly orientated repeat sequences are observed, the disease meets all criteria for a genomic disorder. However, this does not classify the disease as a contiguous gene syndrome. In general, multiple genes are affected by submicroscopic rearrangements, but not all genes may contribute to the phenotype. Thus, extensive genotype-phenotype correlations should be performed to determine whether the observed phenotype can be attributed to dosage variation of a single gene located within the rearranged interval, classifying the disease as a single gene disorder, or whether multiple genes located within the rearranged interval are responsible for distinct clinical features, classifying the disease as a contiguous gene syndrome. In this thesis three disease-related rearrangements are reported that can be classified according to the above-mentioned criteria, i.e, CHARGE syndrome, 17q21.31 microdeletion syndrome and non-recurrent complex rearrangements involving the short arm of chromosome 17.

CHARGE syndrome can be classified as a single gene disorder, since the majority of patients carry a single point mutation in the *CHD7* gene, whereas only in a minority of

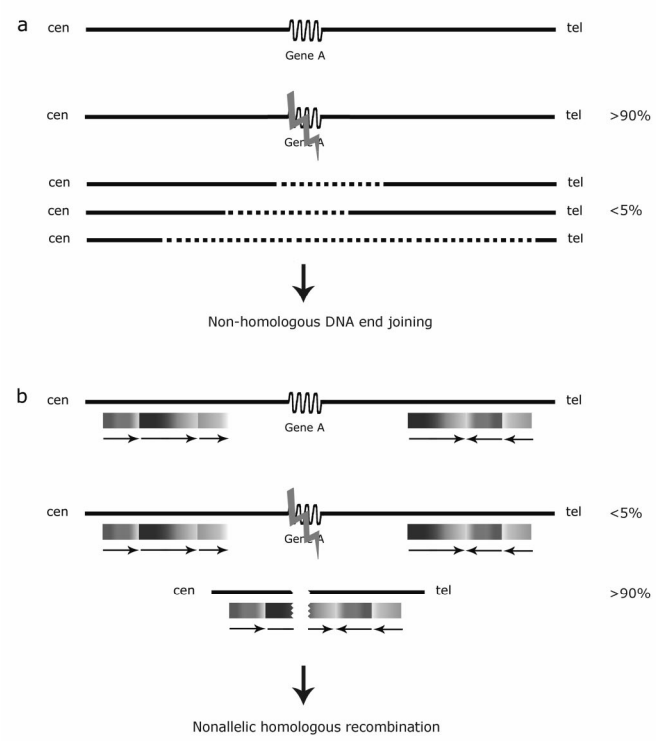


Figure 7.3
Schematic representation of the causes of a single gene disorder and a genomic disorder and the molecular mechanisms involved. The normal genomic orientation is given first with below the mutations leading to disease. (**a**) Single gene disorder such as CHARGE syndrome. The majority of individuals have a mutation in a dosage-sensitive gene, whereas the minority of individuals exhibit a (sub)microscopic deletion. The deletions are all of different size and the respective breakpoints are scattered. All deletions, however, include the dosage-sensitive gene. (**b**) Genomic disorders such as 17q21.31 microdeletion syndrome. The majority of individuals have a submicroscopic deletion caused by nonallelic homologous recombination between the directly orientated LCRs, whereas a minority of individuals has a mutation in a dosage-sensitive gene within the deleted region.

patients the disease is caused by a genomic rearrangement affecting the *CHD7* gene and, in some cases, a few other genes (paragraph 7.3). This notion is supported by several other observations. The severity of the phenotype, for example, does not seem to correlate with the nature of the anomly, i.e., patients with a point mutation in *CHD7* are similarly affected as patients exhibiting chromosomal (micro)deletions involving various additional genes. In addition, the (micro)deletions are non-recurrent and different in size and, thus, exhibit a variable distribution of the breakpoints. Furthermore, detailed *in silico* analyses of the breakpoints of these non-recurrent rearrangements thus far failed to reveal the presence of genomic architectural features such as LCRs (Vissers et al. *unpublished results*). Therefore, other mechanisms, such as NHEJ, have most likely have mediated the formation of the observed (micro)deletions and translocations in CHARGE

syndrome. Based on these data we conclude that CHARGE syndrome is a genuine single gene disorder, sporadically caused by a microdeletion (**Figure 7.3a**).

The second example described in this thesis is the newly discovered 17q21.31 microdeletion syndrome. All three patients, as well as ~25 patients reported by others, exhibited similar sized microdeletions and LCRs flanking the deletions (Chapter 4; Sharp et al. 2006; Shaw-Smith et al. 2006, Koolen et al. *manuscript in preparation*). These characteristics classify the 17q21.31 microdeletion syndrome as a true genomic disorder (**Figure 7.3b**). Given the genomic complexity of the 17q21.31 region, the detection of deletions and inversions, it may only be a matter of time before a (predictable) reciprocal duplication syndrome is identified. Whether the 17q21.31 microdeletion syndrome is a contiguous gene syndrome with sporadic point mutations, or a single gene disorder with sporadic microdeletions, still remains to be determined. The contribution of individual genes located within the deletion interval to the disease phenotype still awaits elucidation. Identification of such mutations may, however, be more difficult than anticipated. In contrast to CHARGE syndrome, the phenotypic spectrum of the 17q21.31 microdeletion syndrome is not fully established yet and, as such, does not allow for a stringent patient selection for subsequent mutation analysis. Additionally, patients with a mutation may present with a completely different phenotype than observed so far for patients exhibiting a microdeletion. Alternatively, there may be no patients with point mutations, and all patients with this syndrome may have microdeletions. However, if a mutated gene exists for this syndrome, the only requirement is that it is dosage-sensitive, thus, leading to haploinsufficiency when mutated. A widely used strategy is to sequence all genes within the deletion interval in patients that fulfill the clinical criteria, but do not carry a microdeletion. Such strategy is relatively easy to accomplish if only a few genes reside within the interval, but may become laborious if many genes reside in it. In the latter case, additional experimental data may provide insight into which gene(s) may serve as candidates, i.e., are dosage-sensitive. For instance, RNA expression-profiling of patients with a microdeletion may reveal decreased expression of one (or few) of the genes located within the deletion interval. Subsequent sequence analysis of such gene(s) in individuals fulfilling the phenotypic criteria, but without microdeletions, may establish whether it is the causative gene for the syndrome under investigation (**Figure 7.4**). Additionally, RNA expression-profiling data may point to dosage-sensitive gene(s) located outside the deletion interval, which may be affected due to a position effect (Merla et al. 2006).

The vast majority of all microdeletions and microduplications represent, however, sporadic, non-recurrent rearrangements. In Chapter 6 examples of two such patients are described, both exhibiting a complex rearrangement involving the short arm of chromosome 17. In total, 17 breakpoints were analyzed of which four were located within LCR structures, suggesting NAHR as the underlying mechanism. The remaining breakpoints were not linked to particular genomic features suggesting that, in addition to NAHR, other mechanisms may have been responsible for the formation of the rearrangements. Although these two examples are not representative for the majority of sporadically occurring rearrangements due to their complexity, they do indicate that NAHR may not be the predominant mechanism in the formation of these rearrangements. Preliminary analysis of 100 chromosomal rearrangements identified in a diagnostic setting, revealed that the majority of the breakpoints did not coincide with currently known genomic architectural features (Koolen et al. *unpublished results*). Thus, as most

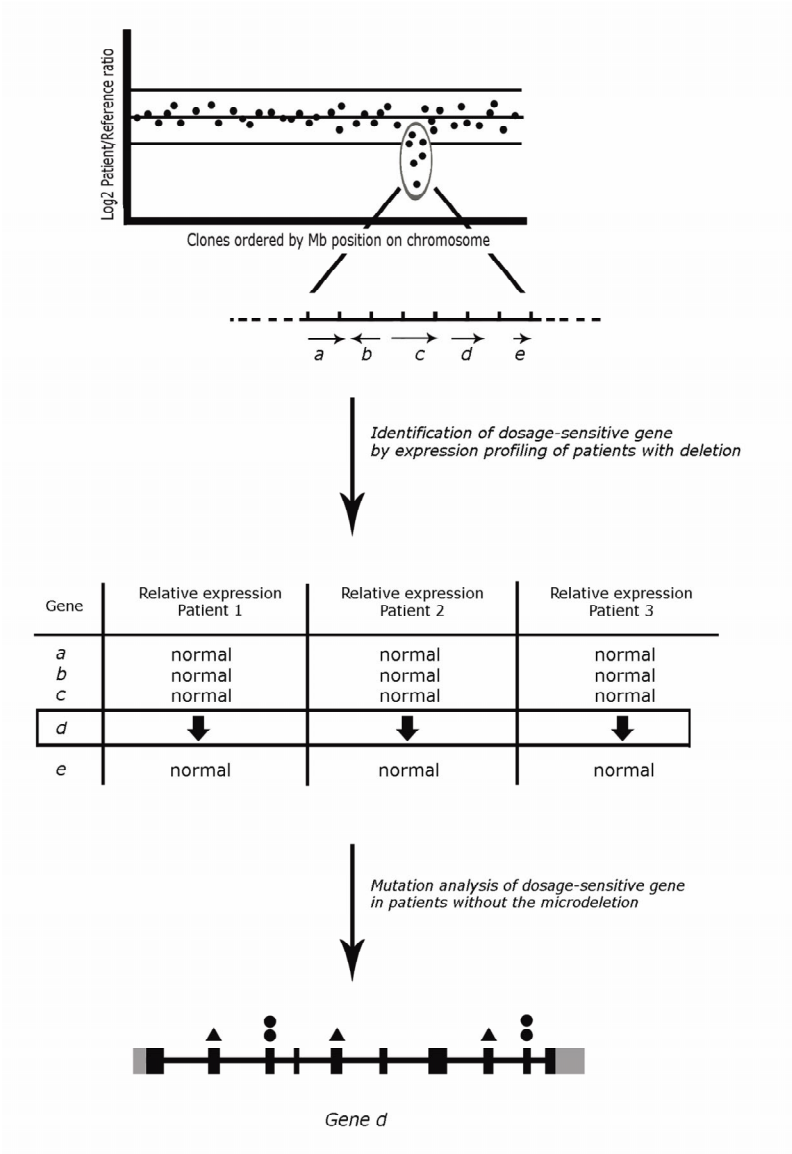


Figure 7.4
Schematic overview of the procedure for finding a dosage-sensitive gene(s). Molecular karyotyping reveals a microdeletion containing genes *'a'* to *'e'* in 3 individual patients. Subsequent expression profiling using RNA from these patients (preferably from a tissue affected by the disease) shows one of the genes, in this example gene *d*, to be decreased in expression relative to a normal control. Sequence analysis of gene *d* in patients with similar phenotypic features but without a microdeletion, may identify mutations in gene *d*, thus explaining the phenotype (in the majority) of the patients.

submicroscopic rearrangements are sporadic and non-recurrent, these observations stress the need to search beyond these architectural features and the mechanisms of NAHR and NHEJ. Alternative architectural features may include common fragile sites, recombination hotspots and/or various recombination-associated motifs. These features are known to predispose to chromosomal rearrangements (Visser et al. 2005; Arlt et al. 2006; Hellenthal & Stephans 2006; Lindsay et al. 2006), but their exact role in the occurrence of sporadic rearrangements remains to be elucidated.

In conclusion, it appears that (syndromic) recurrent rearrangements are mainly caused by NAHR, mediated by the presence of directly orientated LCRs, whereas sporadic, non-recurrent, rearrangements are mainly caused by NHEJ. Since the majority of rearrangements are non-recurrent in nature, future research should be focused on NHEJ (or similar mechanisms) and the genomic architectural features that may mediate these recombination phenomena.

7.6 The complexity of copy number variation

Since the introduction of chromosome banding techniques, there has been awareness of a certain degree of gross genomic variation without clinical significance. Examples of these include variation in heterochromatic centromeric regions, variation in p-arms of acrocentric chromosomes and the occurrence of certain pericentric inversions. In analogy with these cytogenetically visible variations, submicroscopic copy number variations (CNVs) have recently been identified in healthy individuals due to the increased resolution of genome-wide screening technologies such as array CGH. An inventory of such submicroscopic structural CNVs is supported by two databases, i.e., the Database of Genomic Variants (http://projects.tcag.ca/variation/), aiming at cataloging large-scale variation >1kb published in peer-reviewed literature, and the Human Structural Variation Database (http://hgvbase.cgb.ki.se/), a catalogue of human genomic polymorphisms ascertained by experimental and computational analyses. With the publication of a first-generation CNV map of the human genome, it became apparent that CNVs are much more predominant than anticipated until recently. A total of 1447 different CNVs were revealed in 270 healthy individuals (Redon et al. 2006a). These CNVs encompass ~12% of the total human genome (~360 Mb), and can be classified into five different types, being (I) loci showing deletions, (II) loci showing duplications, (III) loci showing both deletions and duplications, (IV) loci showing multi-allelic variation, and (V) complex loci whose precise nature was too difficult to discern. In Chapter 5, a series of 12 genomic loci reported to exhibit CNV in healthy individuals was characterized in detail using MLPA analysis. By doing so, for only seven of these loci CNV was confirmed, whereas based on the initial array CGH data, all regions were classified as type III CNV. In addition, after detailed MLPA analysis, at least three out of the seven loci showed multi-allelic variation, thus classifying these three as type IV CNV. The other four loci remain to be classified as type III CNV. These results underscore the need for (i) validation of CNVs by independent methods and (ii) detailed copy number analyses before cataloging CNVs in databases.

Thousands of functional sequences are affected by loci exhibiting CNV, including genes associated with olfaction, immunity and protein secretion (Nguyen et al. 2006; Redon et al. 2006a). These genes may be amenable to copy number variation as a means to adapt to environmental changes (Zhang et al. 2006). However, there is also increasing evidence

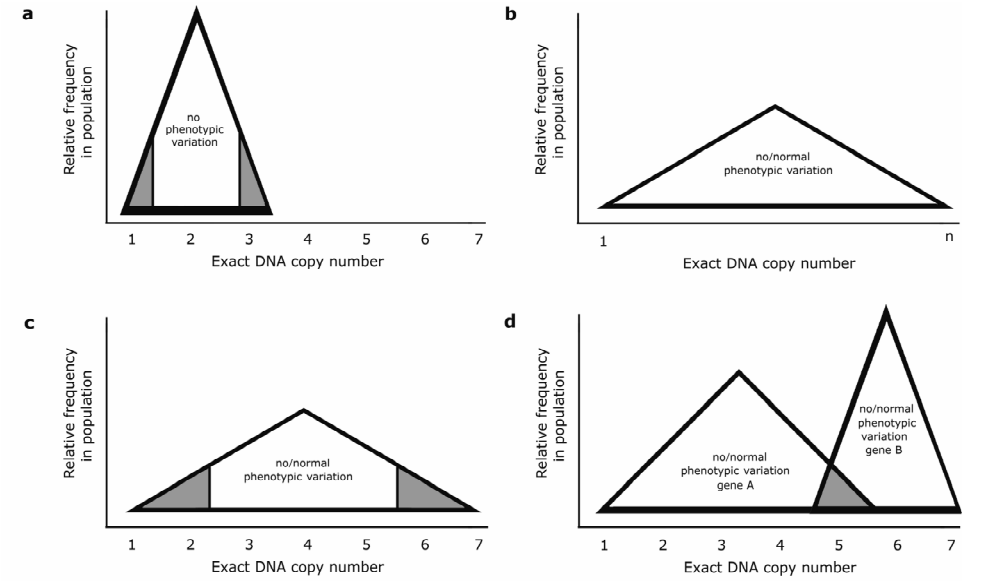


Figure 7.5

Schematic representation of disease-related copy number variation (CNV). (**a**) Old concept: all genes are present in two genomic copies in a healthy person, and deletion or duplication of one copy causes disease. (**b**) New concept: for genes located in CNV loci multiple copies exist without causing disease; instead, they may affect normal phenotypic variation. (**c**) For a subset of such genes, disease may be elicited when the DNA copy number exceeds a certain threshold. (**d**) For complex diseases, two or more genes in multiple CNV loci need to exceed these thresholds in order to cause disease. Threshold levels for disease are indicated by dark gray within the triangle indicating the distribution of DNA copy numbers within the population.

that CNV loci encompass genes that are related to disease. Examples of such genes include the *Fcgr3* gene in renal disease (Aitman et al. 2006) and the *CCL3L1* gene in HIV infection (Gonzalez et al. 2005). The latter two studies have shown that detailed analysis of the exact copy numbers of genes located within such regions is of utmost importance since different copy numbers may have different phenotypic consequences. With this discovery, the dogma that every gene within the genome has two copies and that loss or gain of one such copy may cause disease has become obsolete (**Figure 7.5a**). Instead, the new paradigm is that genes located within CNV loci vary in copy number and that these genes may cause disease once their copy numbers reach above or below certain thresholds (**Figure 7.5b-c**). Additionally, diseases may be caused by a concerted action of two or more genes located in different CNV loci (**Figure 7.5d**). Most likely, linking CNV loci to disease will not be as straightforward as linking a specific gene mutation to a single gene disorder. Instead, linking CNV loci to disease will be similar to linking single nucleotide polymorphisms (SNPs) to common diseases, i.e., the revelation of disease association and disease susceptibility. Related to this notion, it may be expected that some diseases are caused by a combination of CNV loci and SNPs, thus interlinking several levels of normal human variation to disease.

With the new paradigm as described above, the assumption that CNVs are not causative for disease, simply because they are also encountered in healthy individuals, may no longer be justified. Especially in a diagnostic setting this conclusion may have serious consequences, as CNVs inherited through non-affected parents are currently not considered to be causative for disease. This is, however, questionable for a number of reasons apart from those mentioned above. The inherited CNV may, for example, unmask a recessive mutation on the second allele (Lupski & Stankiewicz 2005; Lesnik Oberstein et al. 2006; Marongiu et al. 2007). In order to find such a recessive mutation, all genes located within the deleted region have to be sequenced. This is relatively easy to achieve in case only one inherited copy number imbalance is identified containing a limited number of genes. On average, however, 30-50 CNVs ≥10 kb are encountered, all containing multiple genes, thus making this approach extremely laborious (Conrad et al. 2006; Hinds et al. 2006; McCarroll et al. 2006). Another reason for not rejecting inherited CNVs as causative is that it has been found that affected individuals may inherit a CNV through transmission from an unaffected parent which, instead of being clinically irrelevant, may point at a variable penetrance resulting from dosage variation (Lupski et al. 1991; Chance et al. 1993; Redon et al. 2006a).

In addition to functional genes, 99% of all currently reported CNVs overlap with conserved non-coding sequences (Redon et al. 2006a). Conserved non-coding sequences (CNSs) represent stretches of DNA that are not genic, i.e., they do not produce transcripts (coding or non-coding) with functional properties. There are approximately 327,000 CNSs in the human genome, which together represent 1-2% of this genome. The majority of CNSs are intergenic, thus essentially populating regions not occupied by genes (Dermitzakis et al. 2004; Dermitzakis et al. 2005). Evolutionary analyses strongly support a conserved functional role for these CNSs (Duret et al. 1993; Hardison et al. 1997; Hardison 2000; Glazko et al. 2003; Tanabe et al. 2002) and mutations in CNSs are known to cause disease, such as β-thalassaemia, preaxial polydactyly, and X-linked deafness type 3 (Kioussis et al. 1983; Driscoll et al. 1989; Lettice et al. 2002; Lettice et al. 2003; de Kok et al. 1995b; de Kok et al. 1996). As CNSs may be dosage-sensitive, copy number variation of CNSs may contribute to phenotypic variation and/or disease. Categorizing CNSs located in CNV loci according to their functional characteristics will be a valuable first step in furthering our understanding of the role of CNSs, and their relation to copy number variation in health and disease.

In conclusion, it has undisputedly been acknowledged that large-scale copy number variation within the human genome is much more extensive and complex than anticipated before. Future research should be aimed at studying the genomic content of these CNVs (both genic and non-genic) and a detailed assessment of the copy number spectrum of its sequences. It is anticipated that, in the end, a considerable number of CNVs may be linked to disease, either directly by causing disease, or indirectly by affecting disease susceptibility.

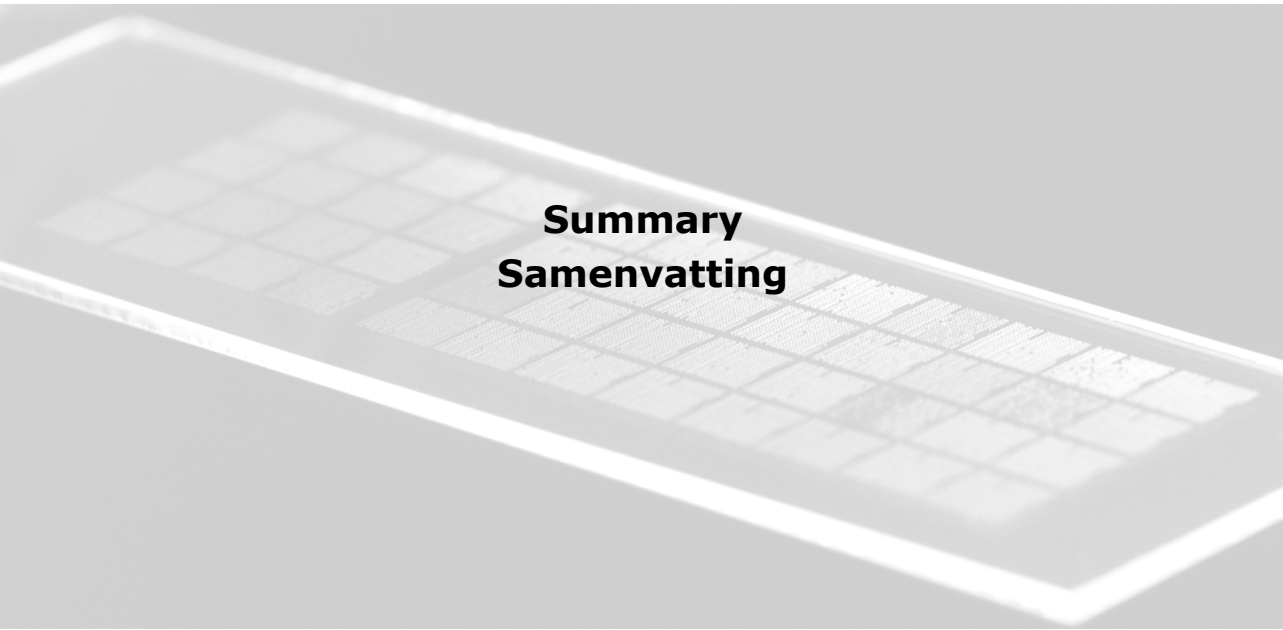
7.7 Concluding remarks

In February 2001, the International Human Genome Sequencing Consortium and Celera Genomics reported the first draft sequence of the human genome (Lander et al. 2001). In the years that followed, this draft sequence has been instrumental for the systematic

analysis of the human genome, including the identification/annotation of novel genes, the elucidation of regional differences in genome composition and the identification of single nucleotide polymorphisms. In addition, new high-throughput approaches such as array CGH were developed that facilitated and markedly accelerated the analysis of the human genome at a large scale, including the detection of an unprecedented level of copy number variation within it. Together, these approaches have contributed significantly to the rapid development of molecular karyotyping, which allows disease phenotypes to be directly linked to gene dosage alterations.

The new concept of molecular karyotyping has significantly changed the field of clinical cytogenetics and clinical diagnostics in the last 4 years and, without doubt, will continue to do so in the years to come. The ability to obtain detailed quantitative copy number information has already led to a significant improvement in diagnostic yield in e.g. patients with mental retardation and is likely to do so for other diseases as well. These diseases will not only include rare congenital disorders but also common complex disorders such as autism, epilepsy, and schizophrenia. The genetic basis of several clinical syndromes has been uncovered by this approach and many more new syndromes await discovery. Finally, detailed insight into human genome plasticity will be obtained by unraveling the genomic architectural features and mechanisms underlying copy number variation in health and disease.

Chapter 8



Summary
Samenvatting

Summary

Chromosomal rearrangements can lead to a wide variety of serious clinical manifestations, including mental retardation and congenital malformation syndromes. Chromosomal rearrangements larger than 5-10 Mb in size can be detected by conventional karyotyping. A considerable number of clinical disorders, however, is caused by submicroscopic chromosomal rearrangements smaller than 5-10 Mb in size. To routinely detect these rearrangements, an efficient and robust genome-wide technology is needed. One such technology is array-based comparative genomic hybridization (array CGH), an approach which is also referred to as molecular karyotyping (Chapter 1). In analogy to the application of array CGH in cancer research, we reasoned that this technology might also be suited for the detection of submicroscopic chromosomal rearrangements in patients with unexplained mental retardation and/or congenital malformation syndromes. The specific aim of this thesis was to explore the feasibility of the array CGH technology for the delineation of DNA copy number alterations for (i) applied clinical genetic and (ii) basic genome research purposes.

Array CGH for the detection of single copy chromosome imbalances in patients with unexplained mental retardation

Mental retardation occurs in 2-3% of the general population. Although a considerable number of cases can be explained by the presence of gross chromosomal rearrangements, in the majority of cases the etiology has so far remained unknown. Submicroscopic chromosomal rearrangements are likely to explain the etiology of mental retardation in at least a proportion of the cases. As outlined above, array CGH may serve as a suitable new technique to capture these rearrangements. In Chapter 2 we described the generation of a microarray consisting of BAC clones selected to cover the entire human genome with a 1 Mb resolution. After optimization, array CGH was employed using this microarray to screen 20 patients with unexplained mental retardation and congenital anomalies for the presence of submicroscopic chromosome rearrangements. Five such rearrangements were identified within this series, two of which appeared to be associated with the clinical features observed. The three remaining rearrangements were considered to be benign copy number variants, as they were also encountered in one of the phenotypically normal parents.

After this pilot study, additional follow-up studies showed that submicroscopic subtelomeric and/or interstitial chromosome rearrangements together explain up to 20% of the previously unexplained cases of mental retardation. Based on these observations we argue that genome-wide copy number profiling technologies, such as array CGH, merit a rapid implementation in the diagnostic testing of patients with unexplained mental retardation (Chapter 7).

Array CGH for resolving the genetic cause of known syndromes

It has been reported that microdeletions may occur at low frequencies in single gene disorders. Based on this information, we hypothesized that the identification of such microdeletions might be instrumental for resolving the genetic causes of clinical disorders with unknown etiologies. In Chapter 3 we tested this hypothesis for CHARGE syndrome, a congenital malformation syndrome denoting a nonrandom pattern of phenotypic features including choanal atresia, malformations of the heart, inner ear and retina. Although several aspects pointed at the involvement of a genetic factor, its cause had remained elusive. Using the genome-wide 1 Mb-resolution BAC array (see above), we identified

overlapping microdeletions on chromosome 8q12 in two patients with CHARGE syndrome. Screening of 17 additional patients with CHARGE syndrome using a tiling resolution chromosome 8 BAC array did not reveal any new microdeletions. Subsequently, all genes residing within the shortest region of deletion overlap were sequenced in these 17 patients. In the majority of them we identified mutations in the *CHD7* gene, thereby resolving the genetic cause of CHARGE syndrome. Subsequent follow-up studies revealed a mutation rate of ~65%, whereby the mutations were scattered throughout the *CHD7* gene without apparent genotype-phenotype correlations. In addition, several (micro)deletions affecting the *CHD7* gene were reported by others, ranging from small exonic deletions to large cytogenetically visible deletions.

By resolving the genetic cause of CHARGE syndrome, we have shown that array CGH can be used as a powerful disease gene identification strategy. As microdeletions and microduplications may comprise up to 15% of all mutations underlying single gene disorders, we anticipate that this strategy will also be applicable to a variety of other clinical disorders in which genetic factors are suspected to be involved (Chapter 7).

Array CGH for unraveling the genetic cause of new syndromes

The possibility to screen large numbers of patients with mental retardation in a high-throughput manner at a genome-wide scale, has significantly increased the chance to identify novel microdeletion and/or microduplication syndromes. Array CGH potentially allows the identification of such novel syndromes through the identification of (partially) overlapping deletions or duplications, without phenotypic preselection of patients. In Chapter 4 we reported the identification of a new microdeletion syndrome involving chromosome region 17q21.31. A total of three overlapping microdeletions within this region were uncovered in a large cohort of patients with mental retardation using high-resolution genome profiling approaches. Detailed clinical examination of these patients revealed a clear phenotypic overlap, including mental retardation, hypotonia and characteristic facial features.

The use of genome-wide profiling technologies does not only facilitate the identification of novel clinical syndromes, but also the establishment of genotype-phenotype correlations. Ultimately, this may lead to the assignment of specific phenotypic traits to individual genes (Chapter 7).

The assessment of large-scale DNA copy number variation within the human genome

Since the introduction of chromosome banding techniques, there has been awareness of a certain degree of genomic variation between individuals, without apparent clinical consequences. The use of high-resolution genome profiling technologies has resulted in the identification of a novel form of genomic variation, i.e., submicroscopic copy number variation (CNV). Such CNVs may encompass large genomic segments ranging in size from a single kilobase to multiple megabases. Due to its recent discovery, the full nature and extent of these CNVs in the human genome is still largely unexplored. In Chapter 5 we described a detailed analysis of twelve CNVs in five different ethnic populations. We characterized the genomic architecture of the chromosomal regions involved and determined individual CNV frequencies in the different ethnic populations. Moreover, we estimated the actual DNA copy numbers of several of the CNVs and used this information to determine the relative distribution of DNA copy numbers. These analyses showed

significant inter- and intra-population differences in CNV distributions and actual DNA copy numbers. Our results indicate that (semi-) quantitative analyses of CNVs are a prerequisite for the definite establishment of their role in health and disease.

This notion was further supported by the observation that CNVs play a role in disease susceptibility when DNA copy numbers surpass certain thresholds. In conjunction with recent reports indicating that ~12% of the human genome exhibits CNV, it is generally accepted now that CNVs are much more common and complex than anticipated before. Although additional studies are required to fully unravel the nature and role of CNVs, it is expected that a considerable number of them will be linked to disease and/or disease susceptibility (Chapter 7).

The role of genomic architecture in mediating copy number variation

As yet, little is known about the etiology of CNVs. It is known, however, that certain genomic architectural features such as low-copy repeats (LCRs) may facilitate CNV formation. Due to extensive sequence homology, LCRs can mediate meiotic misalignment, followed by a molecular mechanism referred to as non-allelic homologous recombination (NAHR). Most clinically recognizable microdeletion syndromes appear to result from LCR-mediated NAHR. As such, the recurrent nature of these CNVs does not represent a random event but, instead, reflects the underlying genomic architectural feature.

The vast majority of disease-causing CNVs identified to date is, however, non-recurrent in nature and, consequently, questions the involvement of genomic architectural features such as LCRs in their etiology. In Chapter 6 two examples of complex, non-recurrent, rearrangements involving chromosome 17p are presented, for which we examined the involvement of genomic architectural features, including LCRs. To this end, we mapped in detail all breakpoints using array CGH and dual color FISH. Our results suggest that LCRs in proximal 17p may have facilitated the formation of at least some of these CNVs and, thus, that these genomic architectural features may also be involved in the generation of non-recurrent CNVs. The observation that not all breakpoints coincided with LCRs suggests the presence of other architectural features and/or molecular mechanisms within the human genome that may facilitate the etiology of CNVs. These features and/or mechanisms still remain to be resolved (chapter 7).

Concluding remarks

Over the last few years, the concept of molecular karyotyping has rapidly changed the field of clinical genetics and genome research and, without any doubt, will continue to do so in the years to come. In case of clinical genetics, we conclude that molecular karyotyping has led to (i) a significant increase in the diagnostic yield in patients with unexplained mental retardation and/or malformation syndromes, (ii) the identification of a gene underlying CHARGE syndrome, and (iii) the identification of a novel microdeletion syndrome involving chromosome 17q21.31. In case of genome research, we conclude that molecular karyotyping has led to novel insight into (i) the occurrence of copy number variation within the human genome, and (ii) recombination mechanisms and genomic architectural features underlying the occurrence of copy number variation. Based on the above, we conclude that molecular karyotyping serves as a powerful tool for linking gene dosage alterations to (disease) phenotypes.

Samenvatting

Chromosoomafwijkingen kunnen tot ernstige klinische aandoeningen leiden zoals verstandelijke handicaps en aangeboren afwijkingen. Chromosoomafwijkingen, groter dan 5-10 Mb, kunnen worden gedetecteerd via conventionele karyotypering. Een aanzienlijk aantal klinische aandoeningen wordt echter veroorzaakt door submicroscopische afwijkingen, kleiner dan 5-10 Mb en zijn derhalve niet detecteerbaar met karyotypering. Om deze kleine afwijkingen routinematig op te kunnen sporen is een nieuwe efficiënte techniek nodig om het genoom te onderzoeken met een resolutie die hoger is dan conventionele karyotypering. Een techniek die daarvoor in aanmerking komt is 'microarray-based comparative genomic hybridization' (array CGH), een methode die ook met de term 'moleculaire karyotypering' wordt aangeduid (Hoofdstuk 1). Op basis van de succesvolle toepassingen van array CGH in het kankeronderzoek, veronderstelden wij dat deze technologie ook toepasbaar zou kunnen zijn voor het opsporen van submicroscopische afwijkingen in patiënten met onverklaarde mentale retardatie (verstandelijke handicap) en/of aangeboren malformatie syndromen. Het specifieke doel van dit proefschrift was dan ook om de toepasbaarheid van array CGH te onderzoeken voor de detectie van (submicroscopische) chromosoomafwijkingen in (i) toegepast klinisch genetisch onderzoek en (ii) fundamenteel genoomonderzoek.

Array CGH voor de detectie van submicroscopische chromosoom afwijkingen in patiënten met onverklaarde mentale retardatie

Mentale retardatie komt in 2-3% van de populatie voor. Hoewel de oorzaak van de mentale retardatie in een deel van de patiënten verklaard wordt door de aanwezigheid van grote chromosoomafwijkingen, wordt in de meerderheid van de patiënten nooit een verklaring gevonden. In een deel van deze patiënten kan de mentale retardatie wellicht verklaard worden door de aanwezigheid van submicroscopische chromosoomafwijkingen. Zoals hierboven beschreven zou array CGH een geschikte techniek kunnen zijn om deze afwijkingen te detecteren. Hoofdstuk 2 beschrijft de ontwikkeling van een microarray die het humane genoom representeert met één BAC kloon per megabase. Na optimalisatie werd deze microarray aangewend om 20 patiënten met onverklaarde mentale retardatie en aangeboren afwijkingen te screenen op de aanwezigheid van submicroscopische chromosoomafwijkingen. In deze serie werden vijf van zulke afwijkingen gevonden, waarvan twee geassocieerd lijken te zijn met de klinische kenmerken van de patiënt. De drie overige afwijkingen werden beschouwd als goedaardige DNA kopieafwijkingen omdat ze tevens werden teruggevonden in één van de niet aangedane ouders.

Na deze eerste studie heeft aanvullend onderzoek laten zien dat submicroscopische subtelomere en/of interstitiële chromosoomafwijkingen tezamen ~20% van alle mentale retardaties verklaart. Gebaseerd op deze cijfers stellen wij voor om technieken die genoomwijde DNA kopieaantallen bepalen, zoals array CGH, op snelle wijze te implementeren in het diagnostisch testen van patiënten met onverklaarde mentale retardatie (Hoofdstuk 7).

Array CGH voor de detectie van de genetische oorzaak van bekende syndromen

Eerder werd gerapporteerd dat microdeleties in een lage frequentie kunnen voorkomen bij monogene aandoeningen. Gebaseerd op deze bevinding postuleerden wij dat het identificeren van dergelijke microdeleties van essentieel belang kan zijn om de genetische oorzaak van klinische aandoeningen met een vooralsnog onbekende etiologie te

achterhalen. In Hoofdstuk 3 hebben we deze hypothese getoetst aan de hand van CHARGE syndroom, een aangeboren aandoening die gekarakteriseerd wordt door een combinatie van fenotypische kenmerken, waaronder afwijkingen van het hart, oor, oog en mentale retardatie. Hoewel verscheidene aspecten wijzen op de betrokkenheid van een genetische factor werd deze nooit opgehelderd. Door gebruik te maken van een 1 Mb resolutie BAC array hebben we in twee patiënten met CHARGE syndroom overlappende microdeleties gevonden in de chromosomale regio 8q12. Het screenen van 17 additionele patiënten met CHARGE syndroom, gebruik makend van een 'tiling resolution' chromosoom 8 array, leidde niet tot de identificatie van extra overlappende microdeleties. Derhalve werd bij deze 17 patiënten de exacte DNA-sequentie bepaald van alle genen die in het minimale gebied van deletieoverlap gesitueerd waren. In de meerderheid van de patiënten werd een mutatie gevonden in het *CHD7* gen, waarmee de genetische oorzaak van CHARGE syndroom opgehelderd werd. Vervolgstudies hebben inmiddels laten zien dat ~65% van de patiënten met CHARGE syndroom een mutatie heeft in het *CHD7* gen en dat de mutaties verspreid zijn over het hele gen, zonder duidelijke genotype-fenotype correlaties. Tevens zijn er additionele (micro)deleties gevonden die het *CHD7* gen bevatten, variërend in grootte van exondeleties tot cytogenetisch zichtbare deleties.

Met de opheldering van de genetische oorzaak van CHARGE syndroom is aangetoond dat array CGH kan fungeren als een krachtige genidentificatie strategie. Geschat wordt dat ~15% van alle monogene aandoeningen veroorzaakt wordt door microdeleties en/of microduplicaties. Op basis hiervan verwachten wij dat deze genidentificatie strategie ook succesvol kan zijn voor een scala van andere syndromen waarvan de genetische oorzaak nog niet opgehelderd is (Hoofdstuk 7).

Array CGH voor het ontrafelen van nieuwe syndromen

Array CGH heeft het mogelijk gemaakt om grote aantallen patiënten met mentale retardatie op genoomwijde schaal te onderzoeken, waardoor de kans om nieuwe syndromen te ontdekken significant verhoogd is. Hoofdstuk 4 rapporteert de ontdekking van een nieuw microdeletie syndroom waarbij de chromosomale regio 17q21.31 betrokken is. In een grote cohort van patiënten met mentale retardatie werden drie overlappende microdeleties geïdentificeerd met behulp van array CGH. Gedetailleerde klinische evaluatie van deze drie patiënten liet, naast mentale retardatie, duidelijke andere overeenkomsten zien, waaronder verlaagde spierspanning en karakteristieke gezichtskenmerken.

Het toepassen van genoomwijde onderzoekstechnieken zoals array CGH vergemakkelijkt niet alleen het ontrafelen van nieuwe syndromen, maar ook genotype-fenotype typeringen. Uiteindelijk kan dit leiden tot het koppelen van specifieke fenotypen aan individuele genen (Hoofdstuk 7).

De bepaling van grootschalige DNA kopieaantal varianten in het menselijk genoom

Sinds de introductie van conventionele karyotypering is bekend dat er een bepaalde mate van genomische variatie bestaat tussen individuen zonder dat er klinische consequenties aan verbonden zijn (normale variatie). Het gebruik van nieuwe genoomwijde technieken heeft naast de ontdekking van ziekte-gerelateerde veranderingen in DNA kopieaantallen geresulteerd in de ontdekking van een nieuwe vorm van normale variatie, namelijk submicroscopische DNA kopieaantal variatie ('CNV'). CNVs kunnen genomische segmenten omvatten die in grootte variëren van één kilobase tot meerdere megabasen.

Omdat CNVs pas recentelijk zijn ontdekt, zijn de verschillende eigenschappen en de totale omvang ervan nog grotendeels onbekend. Hoofdstuk 5 beschrijft de gedetailleerde analyse van twaalf CNVs in vijf etnisch verschillende populaties. We karakteriseerden de genoomarchitectuur van de chromosomale regio's waarin de CNVs voorkomen en bepaalden de frequentie van de CNVs in de verschillende etnische populaties. Daarnaast bepaalden we het daadwerkelijke DNA kopieaantal en gebruikten we deze informatie om de relatieve verdeling van DNA kopieaantallen te bepalen. Deze analyses brachten significante verschillen in CNV verdeling en daadwerkelijke DNA kopieaantallen aan het licht, zowel binnen populaties als tussen populaties. Onze resultaten geven aan dat (semi-) kwantitatieve CNV analyses van essentieel belang zijn voor het definitief vaststellen van hun rol in gezondheid en ziekte.

Deze conclusie wordt ondersteund door de recente bevinding dat CNVs een rol spelen bij de aanleg voor ziekte als het exacte DNA kopieaantal een bepaalde grenswaarde overschrijdt, alsmede door de recente observatie dat 12% van het menselijk genoom onderhevig is aan submicroscopische DNA kopievariatie. Het is nu algemeen geaccepteerd dat CNVs vaker voorkomen en complexer zijn dan voorheen gedacht werd. Hoewel aanvullend onderzoek noodzakelijk is, is het aannemelijk dat een aanzienlijk aantal van deze CNVs uiteindelijk geassocieerd zal kunnen worden met ziekte en/of de aanleg voor ziekte (Hoofdstuk 7).

De rol van genoomarchitectuur in het ontstaan van CNV

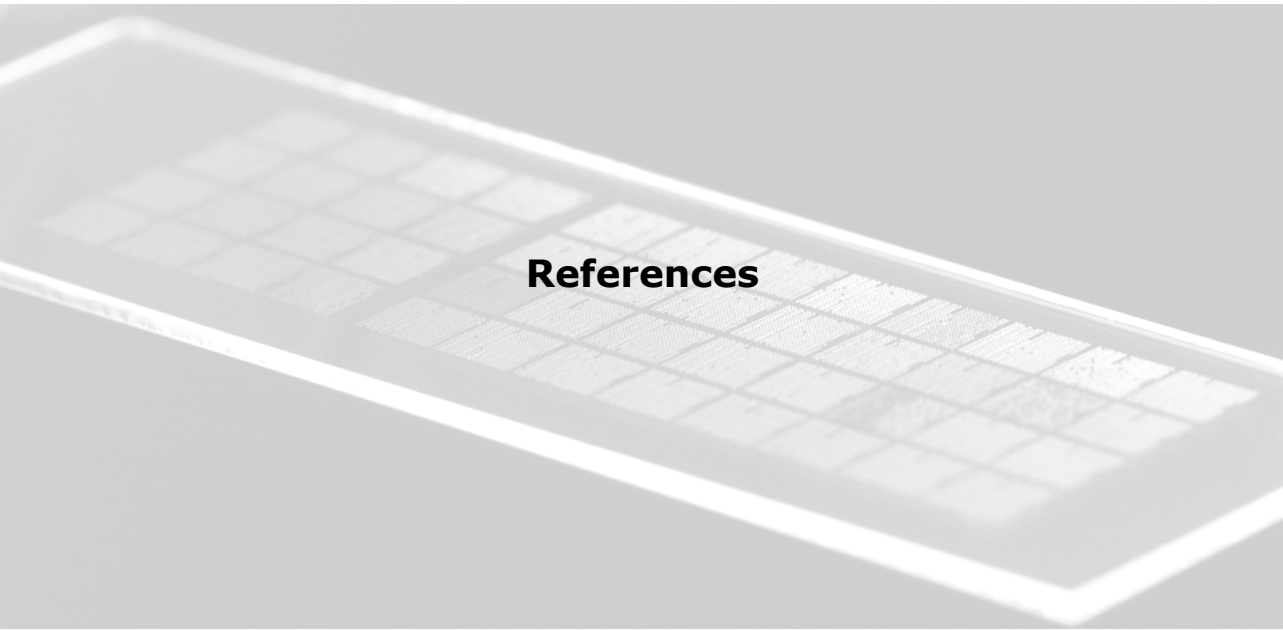
Tot op heden is weinig bekend over de ontstaanswijze van CNVs. Het is echter wel bekend dat bepaalde genomische structuren, zoals 'low-copy repeats' (LCRs), het ontstaan van CNVs kunnen bevorderen. Zo kunnen bijvoorbeeld LCRs, door hun omvangrijke overeenkomsten in sequentie, tijdens de meiose verkeerd rangschikken waardoor een moleculair mechanisme gestimuleerd wordt dat bekend is onder de naam 'nonallelic homologous recombination' (NAHR). De meeste CNVs die ten grondslag liggen aan microdeletie syndromen lijken hun oorsprong te vinden in LCR-gemedieerde NAHR. Hierdoor berust het veelvuldig voorkomen van CNV op een bepaalde genomische locatie niet op toeval, maar op de onderliggende genoomarchitectuur.

De meerderheid van alle CNVs die tot ziekte leidt is echter van sporadische aard en dit plaatst vraagtekens bij de betrokkenheid van genomische structuren zoals LCRs in het ontstaan van deze categorie van CNVs. In Hoofdstuk 6 worden twee voorbeelden gegeven van complexe, sporadische afwijkingen op chromosoom 17p voor welke we de betrokkenheid van deze genomische structuren hebben onderzocht. Daartoe hebben we alle CNV breukpunten gedetailleerd in kaart gebracht met behulp van array CGH en FISH. Onze resultaten suggereren dat LCRs in het proximale deel van 17p het ontstaan van (een deel van) de CNVs gemedieerd hebben. Hiermee is aangetoond dat deze genomische structuren ook betrokken kunnen zijn bij het genereren van sporadische CNVs. De bijkomende observatie dat LCRs niet bij alle breukpunten gevonden werden, suggereert de aanwezigheid van andere genoomstructuren en/of mechanismen die het ontstaan van CNVs kunnen stimuleren. Toekomstig onderzoek zal moeten uitwijzen wat de karakteristieken van dergelijke nieuwe genoomstructuren en/of mechanismen zijn (Hoofdstuk 7).

Tot slot

In de laatste jaren heeft moleculaire karyotypering in razend tempo het vakgebied van de klinische genetica en het genoomonderzoek veroverd. Deze trend zal zich zonder twijfel de komende jaren blijven voorzetten. Voor de klinische genetica heeft de invoering van moleculaire karyotypering geleid tot (i) een significant verhoogde detectie van chromosomale afwijkingen bij patiënten met onverklaarde mentale retardatie en/of aangeboren aandoeningen, (ii) de identificatie van het gen voor CHARGE syndroom, en (iii) de ontdekking van een nieuw 17q21.31 microdeletie syndroom. Met betrekking tot het genoomonderzoek kunnen we concluderen dat moleculaire karyotypering geleid heeft tot nieuwe inzichten in (i) het voorkomen van DNA kopieaantal variaties in het menselijk genoom, en (ii) de recombinatiemechanismen en genoomstructuren die DNA kopieaantal variaties veroorzaken. Samenvattend concluderen we dat moleculaire karyotypering een krachtige strategie gebleken is om veranderingen in gendosis te koppelen aan (ziekte) fenotypen.

Chapter 9



References

Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Roberton-Lowe C, Marshall AJ, Petretto E, et al. *Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans.* Nature **439**: 851-855 (2006).

Albertson DG, Ylstra B, Segraves R, Collins C, Dairkee SH, Kowbel D, Kuo WL, Gray JW, Pinkel D. *Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene.* Nat Genet **25**: 144-156 (2000).

Anderson G, Schroer RJ, Stevenson RE Mental retardation in South Carolina. II. Causation. In: Saul RA, Phelan MC (eds) *Proceedings of the Greenwood Genetic Center*. Greenwood Genetic Center, Greenwood, SC, pp 32–44 (1996).

Aramaki M, Udaka T, Kosaki R, Makita Y, Okamoto N, Yoshihashi H, Oki H, Nanao K, Moriyama N, Oku S, et al. *Phenotypic spectrum of CHARGE syndrome with CHD7 mutations.* J Pediatr **148**: 410-414 (2006).

Aramaki M, Kimura T, Udaka T, Kosaki R, Mitsuhashi T, Okada Y, Takahashi T, Kosaki K. *Embryonic expression profile of chicken CHD7, the ortholog of the causative gene for CHARGE syndrome.* Birth Defects Res Clin Mol Teratol **79**: 50-57 (2007).

Arlt MF, Durkin SG, Ragland RL, Glover TW. *Common fragile sites as targets for chromosome rearrangements*. DNA repair **5**: 1126-1135 (2006).

Arrington CB, Cowley BC, Nightingale DR, Zhou H, Brothman AR, Viskochil DH. *Interstitial deletion 8q11.2-q13 with congenital anomalies of CHARGE association.* Am J Med Genet A **133**: 326-330 (2005).

Axton M. *Structural variants deconstruct the genome*. Nat Genet **38**: 959 (2006).

Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE*. Segmental duplications: organization and impact within the current human genome project assembly.* Genome Res **11**: 1005-1017 (2001).

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. *Recent segmental duplications in the human genome.* Science **297**: 1003-1007 (2002).

Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, Hardy J, Lynch T, Bigio E, Hutton M. *Association of an extended haplotype in the tau gene with progressive supranuclear palsy*. Hum Mol Genet **8**: 711-715 (1999).

Barrett MT, Scheffer A, Ben Dor A, Sampas N, Lipson D, Kincaid R, Tsang P, Curry B, Baird K, Meltzer PS, et al. *Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA.* Proc Natl Acad Sci USA **101**: 17765–17770 (2004).

Bartels I, Starke H, Argyriou L, Sauter SM, Zoll B, Liehr T. *An exceptional complex chromosomal rearrangement (CCR) with eight breakpoints involving four chromosomes (1;3;9;14) in an azoospermic male with normal phenotype.* Eur J Med Genet **50**: 133-138 (2007).

Battisti C, Bonaglia MC, Giglio S, Anichini C, Pucci L, Dotti MT, Zuffardi O, Federico A. *De novo double translocation 3;13 and 4;8;18 in a patient with mental retardation and skeletal abnormalities.* Am J Med Genet A. **117**: 207-211 (2003).

Bayes M, Magano LF, Rivera N, Flores R, Perez Jurado LA*. Mutational mechanisms of Williams-Beuren syndrome deletions.* Am J Hum Genet **73**: 131–151 (2003).

Belloni E, Muenke M, Roessler E, Traverso G, Siegel-Bartelt J, Frumkin A, Mitchell HF, Donis-Keller H, Helms C, et al. *Identification of sonic hedgehog as a candidate gene responsible for holoprosencephaly.* Nat Genet **14**: 353–356 (1996).

Bejjani BA, Saleki R, Ballif BC, Rorem EA, Sundin K, Theisen A, Kashork CD, Shaffer LG. *Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance: is less more?* Am J Med Genet **134**: 259-267 (2005).

Bentley DR. *Whole-genome re-sequencing.* Curr Opin Genet Devel **16**: 545-552 (2006).

Bi W, Yan J, Stankiewicz P, Park SS, Walz K, Boerkoel CF, Potocki L, Shaffer LG, Devriendt K, Nowaczyk MJ, et al. *Genes in a refined Smith-Magenis syndrome critical deletion interval on chromosome 17p11.2 and the syntenic region of the mouse.* Genome Res **12**: 713-728 (2002).

Bi W, Park SS, Shaw CJ, Withers MA, Patel PI, Lupski JR. *Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2.* Am J Hum Genet **73**: 1302-1315 (2003).

Biesecker LG. *The end of the beginning of chromosome ends.* Am J Med Genet **107**: 263–266 (2002).

Bitner-Glindzicz M, Lindley KJ, Rutland P, Blaydon D, Smith VV, Milla PJ, Hussain K, Furth-Lavi J, Cosgrove KE, Shepherd RM, et al. *A recessive contiguous gene deletion causing infantile hyperinsulinism, enteropathy and deafness identifies the Usher type 1C gene.* Nat Genet **26**: 56–60 (2000).

Blake KD, Davenport SL, Hall BD, Hefner MA, Pagon RA, Williams MS, Lin AE, Graham JM Jr. *CHARGE association: an update and review for the primary pediatrician.* Clin Pediatr **37**: 159-173 (1998).

Borg K, Stankiewicz P, Bocian E, Kruczek A, Obersztyn E, Lupski JR, Mazurczak T. *Molecular analysis of a constitutional complex genome rearrangement with 11 breakpoints involving chromosomes 3, 11, 12, and 21 and a approximately 0.5-Mb submicroscopic deletion in a patient with mild mental retardation.* Hum Genet **118**: 267-275 (2005).

Bosman EA, Penn AC, Ambrose JC, Kettleborough R, Stemple DL, Steel K. *Multiple mutations in mouse Chd7 provide models for CHARGE syndrome.* Hum Mol Genet **14**: 3463-3476 (2005).

Brook-Carter PT, Peral B, Ward CJ, Thompson P, Hughes J, Maheshwar MM, Nellist M, Gamble V, Harris PC, Sampson JR. *Deletion of the TSC2 and PKD1 genes associated with severe infantile polycystic kidney disease—a contiguous gene syndrome.* Nat Genet **8**: 328–332 (1994).

Brown SA, Warburton D, Brown LY, Yu CY, Roeder ER, Stengel-Rutkowski S, Hennekam RC, Muenke M. *Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired.* Nat Genet **20**: 180-183 (1998).

Buckley PG, Mantripragada KK, Benetkiewicz M, Tapia-Paez I, Diaz De Stahl T, Rosenquist M, Ali H, Jarbo C, De Bustos C, Hirvela C, et al. *A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications.* Hum Mol Genet **11**: 3221-3229 (2002).

Budarf ML, Emanuel BS. *Progress in the autosomal segmental aneusomy syndromes (SASs): single or multi-locus disorders?* Hum Mol Genet **6**: 1657-1665 (1997).

Caballin MR, Miro R, Egozcue J. *Abnormal phenotype in a child with the same balanced translocation (5;7)(p15;q22) as his father.* Clin Genet **20**: 428–431 (1981).

Cardoso C, Leventer RJ, Ward HL, Toyo-Oka K, Chung J, Gross A, Martin CL, AllansonJ, Pilz DT, Olney AH, et al. *Refinement of a 400-kb critical region allows genotypic*

differentiation between isolated lissencephaly, Miller-Dieker syndrome, and other phenotypes secondary to deletions of 17p13.3. Am J Hum Genet **72**: 918-930 (2003).

Caspersson T, Farber S, Foley GE , Kudynowski J, Modest EJ, Simonsson E, Wagh U, Zech L. *Chemical differentiation along metaphase chromosomes.* Exp Cell Res **49**: 219–222 (1968).

Caspersson T, Zech L, Johansson C. *Differential banding of alkylating fluorochromes in human chromosomes.* Exp Cell Res **60**: 315–319 (1970).

Cavalli G, Paro R.*Chromo-domain proteins: linking chromatin structure to epigenetic regulation.* Curr Opin Cell Biol **10**: 354-360 (1998).

Celli J, van Beusekom E, Hennekam RC, Gallardo ME, Smeets DF, de Cordoba SR, Innis JW, Frydman M, Konig R, Kingston H, et al. *Familial syndromic esophageal atresia maps to 2p23–p24.* Am J Hum Genet **66**: 436–444 (2000).

Chance PF, Alderson MK, Leppig KA, Lensch MW, Matsunami N, Smith B, Swanson PD, Odelberg SJ, Disteche CM, Bird TD*. DNA deletion associated with hereditary neuropathy with liability to pressure palsies.* Cell **72**: 143-151 (1993).

Chen KS, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR. *Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome.* Nat Genet **17**: 154-63 (1997).

Chen CP, Chern SR, Lee CC, Lin CC, Li YC, Hsieh LJ, Chen WL, Wang W *Prenatal diagnosis of de novo t(2;18;14)(q33.1;q12.2;q31.2), dup(5)(q34q34), del(7)(p21.1p21.1), and del(10)(q25.3q25.3) and a review of the prenatally ascertained de novo apparently balanced complex and multiple chromosomal rearrangements.* Prenat Diag **26**: 138-146 (2006).

Cheung VG, Nowak N, Jang W, Kirsch IR, Zhao S, Chen XN, Furey TS, Kim UJ, Kuo WL, Olivier M, et al. *Integration of cytogenetic landmarks into the draft sequence of the human genome*. Nature **409**: 953–958 (2001).

Cheung SW, Shaw CA, Yu W, Li J, Ou Z, Patel A, Yatsenko SA, Cooper ML, Furman P, Stankiewicz P, et al. *Development and validation of a CGH microarray for clinical cytogenetic diagnosis.* Genet Med **7**: 422-432 (2005).

Cleveland WS *Robust locally weighted regression and smoothing scatterplots.* J Amer Stat Assoc **74**: 829–836 (1979).

Cnossen MH, van der Est MN, Breuning MH, van Asperen CJ, Breslau-Siderius EJ, van der Ploeg AT, Goede-Bolder A, van den Ouweland AM, Halley DJ, Niermeijer MF. *Deletions spanning the neurofibromatosis type 1 gene: implications for genotype-phenotype correlations in neurofibromatosis type 1?* Hum Mutat **9**: 458-464 (1997).

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. *A high-resolution survey of deletion polymorphism in the human genome.* Nat Genet **38**: 75-81 (2006).

Crolla JA, van Heyningen V. *Frequent chromosome aberrations revealed by molecular cytogenetic studies in patients with aniridia.* Am J Hum Genet **71**: 1138–1149 (2002).

Cruts M, Rademakers R, Gijselinck I, van der Zee J, Dermaut B, de Pooter T, de Rijk P, Del-Favero J, van Broeckhoven C. *Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region.* Hum Mol Genet **14**: 1753-1762 (2005).

de Baere E, Beysen D, Oley C, Lorenz B, Cocquet J, De Sutter P, Devriendt K, Dixon M, Fellous M, Fryns JP, et al. *FOXL2 and BPES: mutational hotspots, phenotypic variability, and revision of the genotype–phenotype correlation.* Am J Hum Genet **72**: 478–487 (2003).

de Bruijn DR, Kater-Baats E, EleveldM, Merkx G, Geurts Van Kessel A *Mapping and characterization of the mouse and human SS18 genes, two human SS18-like genes and a mouse Ss18 pseudogene*. Cytogenet Cell Genet **92**: 310–319 (2001).

de Kok YJ, van der Maarel SM, Bitner-Glindzicz M, Huber I, Monaco AP, Malcolm S, Pembrey ME, Ropers HH, Cremers FP. *Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4*. Science **267**: 685-688 (1995a).

de Kok YJ, Merkx GF, van der Maarel SM, Huber I, Malcolm S, Ropers HH, Cremers FP. *A duplicateion/paracentric inversion associated with familial X-linked deafness (DFN3) suggests the presence of a regulatory element more than 400 kb upstream of the POU3F4 gene.* Hum Mol Genet **4**: 2145-2150 (1995b).

de Kok YJ, Vossenaar ER, Cremers CW, Dahl N, Laporte J, Hu LJ, Lacombe D, Fischel-Ghodsian N, Friedman RA, Parnes LS, et al. *Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4.* Hum Mol Genet **5**: 1229-1235 (1996).

de Lonlay-Debeney P, Cormier-Daire V, Amiel J, Abadie V, Odent S, Paupe A, Couderc S, Tellier AL, Bonnet D, Prieur M, et al. *Features of DiGeorge syndrome and CHARGE association in five patients.* J Med Genet **34**: 986-989 (1997).

Dermitzakis ET, Kirkness E, Schwarz S, Birney E, Reymond A, Antonarakis SE. *Comparison of human chromosome 21 conserved non-genic sequences (CNGs) with the mouse and dog genomes shows that their selective constraint is independent of their genic environment.* Genome Res **14**: 852-859 (2004).

Dermitzakis ET, Reymond A, Antonarakis SE. *Conserved non-genic sequences – an unexpected feature of mammalian genomes.* Nature Rev Genet **6**: 151-157 (2005).

de Vries BB, van den Ouweland AM, Mohkamsing S, Duivenvoorden HJ, Mol E, Gelsema K, van Rijn M, Halley DJ, Sandkuijl LA, Oostra BA, et al. *Screening and diagnosis for the fragile X syndrome among the mentally retarded: an epidemiological and psychological survey.* Am J Hum Genet **61**: 660–667 (1997).

de Vries BB, White SM, Knight SJ, Regan R, Homfray T, Young ID, Super M, McKeown C, Splitt M, Quarrell OW, et al. *Clinical studies on submicroscopic subtelomeric rearrangements: a checklist.* J Med Genet **38**: 145–150 (2001).

de Vries BB, Winter R, Schinzel A, Van Ravenswaaij-Arts C. *Telomeres: a diagnosis at the end of the chromosomes.* J Med Genet **40**: 385–398 (2003).

de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM, Reijmersdal S, Nillesen WM, Huys EH, Leeuw N, et al. *Diagnostic genome profiling in mental retardation.* Am J Hum Genet **77**: 606-616 (2005).

Dhami P, Coffey AJ, Abbs S, Vermeesch JR, Dumanski JP, Woodward KJ, Andrews RM, Langford C, Vetrie D. *Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome.* Am J HumGenet **76**: 750–762 (2005).

Dobyns WB, Reiner O, Carrozzo R, Ledbetter DH. *Lissencephaly. A human brain malformation associated with deletion of the LIS1 gene located at chromosome 17p13.* JAMA **270:** 2838-2842 (1993).

Donnai D, Karmiloff-Smith A. *Williams syndrome: from genotype through to the cognitive phenotype.* Am J Med Genet **97**: 164-171 (2000).

Douglas J, Hanks S, Temple IK, Davies S, Murray A, Upadhyaya M, Tomkins S, Hughes HE, Cole TR, Rahman N. *NSD1 mutations are the major cause of Sotos syndrome and*

occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes. Am J Hum Genet **72**: 132-143 (2003).

Driscoll MC, Dobkin CS, Alter BP. *γ-δ-β-Thalassaemia due to a de novo mutation deleting the 5′ β-globin gene activation-region hypersensitive sites.* Proc Natl Acad Sci USA **86**: 7470-7474 (1989).

Duret L, Dorkeld F, Gautier C. *Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression.* Nucleic Acids Res **21**: 2315-2322 (1993).

Eichler EE. *Recent duplication, domain accretion and the dynamic mutation of the human genome*. Trends Genet **17**: 661-669 (2001).

Eichler EE. *Widening the spectrum of human genetic variation.* Nat Genet **38**: 9-11 (2006).

Ekong R, Jeremiah S, Judah D, Lehmann O, Mirzayans F, Hung YC, Walter MA, Bhattacharya S, Gant TW, Povey S, Wolfe J. *Chromosomal anomalies on 6p25 in iris hypoplasia and Axenfeld-Rieger syndrome patients defined on a purpose-built genomic microarray.* Hum Mutat **24**: 76-85 (2004).

Eronen M, Peippo M, Raatikka M, Arvio M, Johansson R, Kahkonen M. *Cardiovascular manifestations in 75 patients with Williams syndrome*. J Med Genet **39**: 554–558 (2002).

Feenstra I, Fang J, Koolen DA, Siezen A, Evans C, Winter RM, Lees MM, Riegel M, de Vries BB, van Ravenswaaij CM et al. *European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA); an online database for rare chromosome abnormalities.* Eur J Med Genet **49**: 279-291 (2006).

Feuk L, Carson AR, Scherer SW. *Structural variation in the human genome.* Nat Rev Genet **7**: 85-97 (2006).

Flint J, Wilkie AO, Buckle VJ, Winter RM, Holland AJ, McDermid HE. *The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation.* Nat Genet **9**: 132–140 (1995).

Forozan F, Karhu R, Kononen J, Kallioniemi A, Kallioniemi OP. *Genome screening by comparative genomic hybridization.* Trends Genet **13**: 405–409 (1997).

Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ. *Complex SNP-related sequence variation in segmental genome duplications.* Nat Genet **36**: 861-866 (2004).

Friedman JM, Baross A, Delaney AD, Ally A, Arbour L, Armstrong L, Asano J, Bailey DK, Barber S, Birch P, et al. *Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation.* Am J Hum Genet **79**: 500-513 (2006).

Frydman M, Steinberger J, Shabtai F, Steinherz R. *Interstitial 7q deletion [46,XY,del(7)(pterrcen::q112rqter)] in a retarded quadriplegic boy with normal beta glucuronidase.* Am J Med Genet **25**: 245–249 (1986).

Gajecka M, Glotzbach CD, Shaffer LG. *Characterization of a complex rearrangement with interstitial deletions and inversion on human chromosome 1.* Chromosome Res **14**: 277-282 (2006).

Gimelli G, Pujana MA, Patricelli MG, Russo S, Giardino D, Larizza L, Cheung J, Armengol L, Schinzel A, Estivill X, et al. *Genomic inversions of human chromosome 15q11-q13 in mothers of Angelman syndrome patients with class II (BP2/3) deletions.* Hum Mol Genet **12**: 849-858 (2003).

Glazko GV, Koonin EV, Rogozin IB, Shabalina SA. *A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions.* Trends Genet **19**: 119-124 (2003).

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ et al. *The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility.* Science **307**: 1434-1440 (2005).

Gripp KW, Wotton D, Edwards MC, Roessler E, Ades L, Meinecke P, Richieri-Costa A, Zackai EH, Massague J, Muenke M, Elledge SJ. *Mutations in TGIF cause holoprosencephaly and link NODAL signalling to human neural axis determination.* Nat Genet **25:** 205-208 (2000).

Harada N, Hatchwell E, Okamoto N, Tsukahara M, Kurosawa K, Kawame H, Kondoh T, Ohashi H, Tsukino R, Kondoh Y, et al. *Subtelomere specific microarray based comparative genomic hybridisation: a rapid detection system for cryptic rearrangements in idiopathic mental retardation.* J Med Genet **41**: 130-136 (2004).

Hardison RC, Celtjen J, Miller W. *Long human-mouse sequence alignments reveal novel regulatory elements : a reason to sequence to mouse genome.* Genome Res **7**: 959-966 (1997).

Hardison RC. *Conserved noncoding sequences are reliable guides to regulatory elements.* Trends Genet **16**: 369-372 (2000).

Hellenthal G, Stephans M. *Insights into recombination from population genetic variation.* Curr Opin Genet Dev **16**: 565-572 (2006).

Herr A, Grutzmann R, Matthaei A, Artelt J, Schrock E, Rump A, Pilarsky C. *High-resolution analysis of chromosomal imbalances using the Affymetrix 10K SNP genotyping chip*. Genomics **85**: 392–400 (2005).

Herrera L, Kakati S, Gibas L, Pietrzak E, Sandberg AA. *Gardner syndrome in a man with an interstitial deletion of 5q.* Am J Med Genet **25**: 473-476 (1986).

Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA. *Common deletions and SNPs are in linkage disequilibrium in the human genome.* Nat Genet. **38**: 82-85 (2006).

Hollox EJ, Armour JA, Barber JC. *Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster*. Am J Hum Genet **73**: 591-600 (2003).

Hurst JA, Meinecke P, Baraitser M. *Balanced t(6;8)(6p8p;6q8q) and the CHARGE association.* J Med Genet **28**: 54-55 (1991).

Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, Chakraverty S, Isaacs A, Grover A, et al. *Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17.* Nature **393**: 702-705 (1998).

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. *Detection of large-scale variation in the human genome*. Nat Genet **36**: 949–951 (2004).

Ikegami S, Harada A, Hirokawa N. *Muscle weakness, hyperactivity, and impairment in fear conditioning in tau-deficient mice.* Neurosci Lett **279**: 129-132 (2000).

International HapMap consortium. *A haplotype map of the human genome.* Nature **437**: 1299-1320 (2005).

Inoue K, Dewar K, Katsanis N, Reiter LT, Lander ES, Devon KL, Wyman DW, Lupski JR, Birren B. *The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes.* Genome Res **11**: 1018-1033 (2001).

Inoue K, Lupski JR. *Molecular mechanisms for genomic disorders.* Annu Rev Genomics Hum Genet **3**: 199-242 (2002).

Inoue K, Osaka H, Thurston VC, Clarke JT, Yoneyama A, Rosenbarker L, Bird TD, Hodes ME, Shaffer LG**,** Lupski JR. *Genomic rearrangements resulting in PLP1 deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females.* Am J Hum Genet **71**: 838-853 (2002).

Ion A, Crosby AH, Kremer H, Kenmochi N, van Reen M, Fenske C, van de Burgt I, Brunner HG, Montgomery K, Kucherlapati RS et al. *Detailed mapping, mutation analysis, and intragenic polymorphism identification in candidate Noonan syndrome genes MYL2, DCN, EPS8, and RPL6.* J Med Genet **37**: 884–886 (2000).

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, et al. *A tiling resolution DNA microarray with complete coverage of the human genome.* Nat Genet. **36**: 299-303 (2004).

Jacobs PA, Hunt PA, Mayer M, Bart RD. *Duchenne muscular dystrophy (DMD) in a female with an X/autosome translocation: further evidence that the DMD locus is at Xp21.* Am J Hum Genet **33**: 513-518 (1981).

Johnson D, Horsley SW, Moloney DM, Oldridge M, Twigg SR, Walsh S, Barrow M, Njolstad PR, Kunz J, Ashworth GJ, et al. *A comprehensive screen for TWIST mutations in patients with craniosynostosis identifies a new microdeletion syndrome of chromosome band 7p21.1.* Am J Hum Genet **63**: 1282-1293 (1998).

Johnson D, Morrison N, Grant L, Turner T, Fantes J, Connor JM, Murday V. *Confirmation of CHD7 as a cause of CHARGE association identified by mapping a balanced chromosome translocation in affected monozygotic twins.* J Med Genet **43**: 280-284 (2006).

Johnston JJ, Walker RL, Davis S, Facio F, Turner JT, Bick DP, Daentl DL, Ellison JW, Meltzer PS, Biesecker LG. *Zoom-in comparative genomic hybridisation arrays for the characterisation of variable breakpoint contiguous gene syndromes.* J Med Genet **44**: e59 (2007).

Jongmans MC, Admiraal RJ, van der Donk KP, Vissers LE, Baas AF, Kapusta L, van Hagen JM, Donnai D, de Ravel TJ, Veltman JA, et al. *CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene.* J Med Genet **43**: 306-314 (2006).

Juyal RC, Figuera LE, Hauge X, Elsea SH, Lupski JR, Greenberg F, Baldini A, Patel PI. *Molecular analyses of 17p11.2 deletions in 62 Smith-Magenis syndrome patients.* Am J Hum Genet **58**: 998-1007 (1996).

Kallen K, Robert E, Mastroiacovo P, Castilla EE, Kallen B. *CHARGE Association in newborns: a registry-based study.* Teratology **60**: 334-343 (1999).

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. *Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors.* Science **258**: 818-821 (1992).

Karmous-Benailly H, Giuliano F, Massol C, Bloch C, De Ricaud D, Lambert JC, Perelman S. *Unbalanced inherited complex chromosome rearrangement involving chromosome 8, 10, 11 and 16 in a patient with congenital malformations and delayed development.* Eur J Med Genet **49**: 431-438 (2006).

Kent WJ. *BLAT--the BLAST-like alignment tool.* Genome Res **12**: 656-664 (2002).

Kioussis D, Vanin E, deLange T, Flavell RA, Grosveld FG. *β-globin gene inactivation by DNA traslocation un γ-β-thalassaemia.* Nature **306**: 662-666 (1983).

Kishino T, Lalande M, Wagstaff J. *UBE3A/E6-AP mutations cause Angelman syndrome.* Nat Genet **15**: 70-73 (1997).

Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmuller G. *Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells.* Proc Natl Acad Sci USA **96**: 4494-4499 (1999).

Klein OD, Cotter PD, Albertson DG, Pinkel D, Tidyman WE, Moore MW, Rauen KA. *Prader-Willi syndrome resulting from an unbalanced translocation: characterization by array comparative genomic hybridization.* Clin Genet **65**: 477-482 (2004).

Knight SJ, Regan R, Nicod A, Horsley SW, Kearney L, Homfray T, Winter RM, Bolton P, Flint J. *Subtle chromosomal rearrangements in children with unexplained mental retardation.* Lancet **354**: 1676–1681 (1999).

Knight SJ, Flint J. *Perfect endings: a review of subtelomeric probes and their use in clinical diagnosis.* J Med Genet **37**: 401-409 (2000).

Knight SJ, Lese CM, Precht KS, Kuc J, Ning Y, Lucas S, Regan R, Brenan M, Nicod A, Lawrie NM, et al. *An optimized set of human telomere clones for studying telomere integrity and architecture.* Am J Hum Genet **67**: 320–332 (2000).

Knight SJ, Regan R. *Idiopathic learning disability and genome imbalance.* Cytogenet Genome Res **115**: 215-224 (2006).

Koolen DA, Nillesen WM, Versteeg MH, Merkx GF, Knoers NV, Kets M, Vermeer S, van Ravenswaaij CM, de Kovel CG, Brunner HG, et al. *Screening for subtelomeric rearrangements in 210 patients with unexplained mental retardation using multiplex ligation dependent probe amplification (MLPA).* J Med Genet **41**: 892-899 (2004).

Koolen DA, Reardon W, Rosser EM, Lacombe D, Hurst JA, Law CJ, Bongers EM, van Ravenswaaij-Arts CM, Leisink MA, Geurts van Kessel A, et al. *Molecular characterization of patients with subtelomeric 22q abnormalities using chromosome specific array-based comparative genomic hybridisation.* Eur J Hum Genet **13**: 1019-1024 (2005).

Koppens PF, Hoogenboezem T, Degenhart HJ. *CYP21 and CYP21P variability in steroid 21-hydroxylase deficiency patients and in the general population in the Netherlands.* Eur J Hum Genet **8**: 827–836 (2000).

Krantz ID, Rand EB, Genin A, Hunt P, Jones M, Louis AA, Graham JM Jr., Bhatt S, Piccoli DA, Spinner NB*. Deletions of 20p12 in Alagille syndrome: frequency and molecular characterization.* Am J Med Genet **70**: 80-86 (1997).

Kriek M, White SJ, Szuhai K, Knijnenburg J, van Ommen GJ, den Dunnen JT, Breuning MH. *Copy number variation in regions flanked (or unflanked) by duplicons among patients with developmental delay and/or congenital malformations; detection of reciprocal and partial Williams-Beuren duplications.* Eur J Hum Genet **14**: 180-189 (2006).

Kuechler A, Ziegler M, Blank C, Rommel B, Bullerdiek J, Ahrens J, Claussen U, Liehr T. *A highly complex chromosomal rearrangement between five chromosomes in a healthy female diagnosed in preparation for intracytoplasmatic sperm injection.* J Histochem Cytochem **53**: 355-357 (2005).

Kurotaki N, Harada N, Shimokawa O, Miyake N, Kawame H, Uetake K, Makita Y, Kondoh T, Ogata T, Hasegawa T, et al. *Fifty microdeletions among 112 cases of Sotos syndrome: low copy repeats possibly mediate the common deletion.* Hum Mutat **22**: 378-387 (2003).

Lalani SR, Stockton DW, Bacino C, Molinari LM, Glass NL, Fernbach SD, Towbin JA, Craigen WJ, Graham JM Jr, Hefner MA, et al. *Toward a genetic etiology of CHARGE syndrome:*

I. A systematic scan for submicroscopic deletions. Am J Med Genet A **118**: 260-266 (2003).

Lalani SR, Safiullah AM, Fernbach SD, Harutyunyan KG, Thaller C, Peterson LE, McPherson JD, Gibbs RA, White LD, Hefner M, et al. *Spectrum of CHD7 Mutations in 110 Individuals with CHARGE Syndrome and Genotype-Phenotype Correlation.* Am J Hum Genet **78**: 303-314 (2006).

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. *International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome.* Nature **409**: 860-921 (2001).

Langdon JA, Lamont JM, Scott DK, Dyer S, Prebble E, Bown N, Grundy RG, Ellison DW, Clifford SC. *Combined genome-wide allelotyping and copy number analysis identify frequent genetic losses without copy number reduction in medulloblastoma.* Genes Chromosomes Cancer **45**: 47-60 (2006).

Larrabee PB, Johnson KL, Pestova E, Lucas M, Wilber K, LeShane ES, Tantravahi U, Cowan JM, Bianchi DW. *Microarray analysis of cell-free fetal DNA in amniotic fluid: a prenatal molecular karyotype.* Am J Hum Genet **75**: 485-491 (2004).

Lee JA, Inoue K, Cheung SW, Shaw CA, Stankiewicz P, Lupski JR. *Role of genomic architecture in PLP1 duplication causing Pelizaeus-Merzbacher disease.* Hum Mol Genet **15**: 2250-2265 (2006).

Lele KP, Penrose LS, Stallard HB. *Chromosome deletion in a case of retinoblastoma.* Ann Hum Genet **27**: 171-174 (1963).

Lesnik Oberstein, Kriek M, White SJ, Kalf ME, Szuhai K, den Dunnen JT, Breuning MH, Hennekam RC. *Peters Plus syndrome is caused by mutations in B3GALTL, a putative glycosyl transferase.* Am J Hum Genet **79**: 562-566 (2006).

Lespinasse J, Bugge M, Rethore MO, North MO, Lundsteen C, Kirchhoff M. *De novo complex chromosomal rearrangements (CCR) involving chromosome 1, 5, and 6 resulting in microdeletion for 6q14 in a female carrier with psychotic disorder.* Am J Med Genet A **128**: 199-203 (2004).

Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, et al. *Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly.* Proc Natl Acad Sci USA **99**: 7548-7553 (2002).

Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. *A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly.* Hum Mol Genet **12**: 1725-1735 (2003).

Lichter P. *New tools in molecular pathology.* J Mol Diagn **2**: 171-173 (2000).

Lieber MR, Yu K, Raghavan SC. *Roles of nonhomologous DNA end joining, V(D)J recombination, and class switch recombination in chromosomal translocations.* DNA repair **5**: 1234-1245 (2006).

Liehr T, Starke H, Weise A, Lehrer H, Claussen U. *Multicolor FISH probe sets and their applications.* Histol Histopathol **19**: 229-237 (2004).

Lindsay EA. *Chromosomal microdeletions: dissecting del22q11 syndrome.* Nat Rev Genet **2**: 858-868 (2001).

Lindsay EA, Vitelli F, Su H, Morishima M, Huynh T, Pramparo T, Jurecic V, Ogunrinu G, Sutherland HF, Scambler PJ, Bradley A, Baldini A. *Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice.* Nature **410**: 97-101 (2001).

Lindsay SJ, Khajavi M, Lupski JR, Hurles ME. *A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination.* Am J Hum Genet **79**: 890-902 (2006).

Locke DP, Segraves R, Nicholls RD, Schwartz S, Pinkel D, Albertson DG, Eichler EE. *BAC microarray analysis of 15q11-q13 rearrangements and the impact of segmental duplications.* J Med Genet **41**: 175-182 (2004).

Luijendijk MW, van de Pol TJ, van Duijnhoven G, den Hollander AI, ten Caat J, van Limpt V, Brunner HG, Kremer H, Cremers FP. *Cloning, characterization, and mRNA expression analysis of novel human fetal cochlear cDNAs.* Genomics **82**: 480-490 (2003).

Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask BJ, Saucedo-Cardenas O, Barker DF, Killian JM, Garcia CA, et al. *DNA duplication associated with Charcot-Marie-Tooth disease type 1A*. Cell **66**: 219-232 (1991).

Lupski JR. *Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits.* Trends Genet **14**: 417-422 (1998).

Lupski JR, Stankiewicz P. *Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes.* PLoS Genet **1**: e49 (2005).

Lupski JR. *Genome structural variation and sporadic disease traits.* Nat Genet **38**:974-976 (2006).

Lupski JR, Stankiewicz P. Editors of: *Genomic Disorders: the genomic basis of disease.* Humana Press, Totowa, New Jersey (2006).

Marongiu R, Brancati F, Antonini A, Ialongo T, Ceccarini C, Scarciolla O, Capalbo A, Benti R, Pezzoli G, Dallapiccola B, et al. *Whole gene deletion and splicing mutations expand the PINK1 genotypic spectrum.* Hum Mutat **28**: 98-106 (2007).

Martin DM, Probst FJ, Fox SE, Schimmenti LA, Semina EV, Hefner MA, Belmont JW, Camper SA. *Exclusion of PITX2 mutations as a major cause of CHARGE association.* Am J Med Genet **111**: 27-30 (2002).

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly M et al. *Common deletion polymorphisms in the human genome.* Nat Genet **38**: 86-92 (2006).

Menten B, Maas N, Thienpont B, Buysse K, Vandesompele J, Melotte C, de Ravel T, Van Vooren S, Balikova I, Backx L, et al. *Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports.* J Med Genet **43**: 625-633 (2006).

Menten B, Buysse J, Zahir F, Hellemans J, Hamilton SJ, Costa T, Fagerstrom C, Anadiotis G, Kingsbury D, McGillivray BC et al. *Osteopoikilosis, short stature, and mental retardation as key features of a new microdeletion syndrome on 12q14.* J Med Genet **44**: 264-268 (2007).

Merla G, Howald C, Henrichsen CN, Lyle R, Wyss C, Zabot MT, Antonarakis SE, Reymond A. *Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes.* Am J Hum Genet **79**: 332-341 (2006).

Mimault C, Giraud G, Courtois V, Cailloux F, Boire JY, Dastugue B, Boespflug-Tanguy O. *Proteolipoprotein gene analysis in 82 patients with sporadic Pelizaeus-Merzbacher Disease: duplications, the major cause of the disease, originate more frequently in*

male germ cells, but point mutations do not. The Clinical European Network on Brain Dysmyelinating Disease. Am J Hum Genet **65**: 360-369 (1999).

Mizugishi K, Yamanaka K, Kuwajima K, Kondo I. *Interstitial deletion of chromosome 7q in a patient with Williams syndrome and infantile spasms.* J Hum Genet **43**: 178–181 (1998).

Mowat DR, Wilson MJ, Goossens M. *Mowat-Wilson syndrome.* J Med Genet **40**: 305–310 (2003).

Mueller RF, Young ID. *Emery's elements of medical genetics.* 11th edition, page 35-53 (2001).

Munke M. *Clinical, cytogenetic, and molecular approaches to the genetic heterogeneity of holoprosencephaly.* Am J Med Genet **34**: 237-245 (1989).

Nagase T, Kikuno R, Ishikawa KI, Hirosawa M, Ohara O. *Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro.* DNA Res **7**: 65-73 (2000).

Nguyen DQ, Webber C, Ponting CP. *Bias of selection on human copy-number variants.* PLoS Genet **2**: e20 (2006).

Olney RC, Mougey EB, Wang J, Shulman DI, Sylvester J.E. *Using real-time, quantitative PCR for rapid genotyping of the steroid 21-hydroxylase gene in a north Florida population.* J Clin Endocrinol Metab **87**: 735-741 (2002).

Onufer CN, Stephan MJ, Thuline HC, Char F. *Chromosome 13 long arm interstitial deletion associated with features of Noonan phenotype*. Ann Genet **30**: 236–239 (1987).

Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, Costa T, Grebe T, Cox S, Tsui LC, et al. *A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome.* Nat Genet **29**: 321-325 (2001).

Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, de Jong PJ. *A bacterial artificial chromosome library for sequencing the complete human genome.* Genome Res **11**: 483–496 (2001).

Ou Z, Jarmuz M, Sparagana SP, Michaud J, Decarie JC, Yatsenko SA, Nowakowska B, Furman P, Shaw CA, Shaffer LG, et al. *Evidence for involvement of TRE-2 (USP6) oncogene, low-copy repeat and acrocentric heterochromatin in two families with chromosomal translocations.* Hum Genet **120**: 227-237 (2006).

Pagon RA, Graham JM Jr, Zonana J, Yong SL. *Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association.* J Pediatr **99**: 223-237 (1981).

Pai GS, Thomas GH, Mahoney W, Migeon BR. *Complex chromosome rearrangements. Report of a new case and literature review.* Clin Genet **18**: 436-444 (1980).

Parvari R, Brodyansky I, Elpeleg O, Moses S, Landau D, Hershkovitz E. *A recessive contiguous gene deletion of chromosome 2p16 associated with cystinuria and a mitochondrial disease.* Am J Hum Genet **69**: 869–875 (2001).

Patel PI, Roa BB, Welcher AA, Schoener-Scott R, Trask BJ, Pentao L, Snipes GJ, Garcia CA, Francke U, Shooter EM, et al. *The gene for the peripheral myelin protein PMP-22 is a candidate for Charcot-Marie-Tooth disease type 1A.* Nat Genet **1**: 159-165 (1992).

Patsalis PC, Evangelidou P, Charalambous S, Sismani C. *Fluorescence in situ hybridization characterization of apparently balanced translocation reveals cryptic complex chromosomal rearrangements with unexpected level of complexity.* Eur J Hum Genet **12**: 647-653 (2004).

Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, Haden K, Li J, Shaw CA, Belmont J, et al. *High-resolution profiling of chromosome aberrations using Infinium whole-genome genotyping.* Genome Res **16**: 1136-1148 (2006).

Pentao L, Wise CA, Chinault AC, Patel PI, Lupski JR. *Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit.* Nat Genet **2**: 292-300 (1992).

Petrij F, Dauwerse HG, Blough RI, Giles RH, van der Smagt JJ, Wallerstein R, Maaswinkel-Mooy PD, van Karnebeek CD, van Ommen GJ, van Haeringen A, et al. *Diagnostic analysis of the Rubinstein-Taybi syndrome: five cosmids should be used for microdeletion detection and low number of protein truncating mutations.* J Med Genet **37**: 168-176 (2000).

Pfeifer D, Pantic M, Skatulla I, Rawluk J, Kreutz C, Martens UM, Fisch P, Timmer J, Veelken H. *Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays.* Blood **109**: 1202-1210 (2007).

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, et al. *High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.* Nat Genet **20**: 207-211 (1998).

Pinkel D, Albertson DG. *Array comparative genomic hybridization and its applications in cancer.* Nat Genet **37** Suppl: S11-17 (2005).

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. *Genome-wide analysis of DNA copy-number changes using cDNA microarrays.* Nat Genet **23**: 41–46 (1999).

Potocki L, Shaffer LG. *Interstitial deletion of 11(p11.2p12): a newly described contiguous gene deletion syndrome involving the gene for hereditary multiple exostoses (EXT2).* Am J Med Genet **62**: 319–325 (1996).

Potocki L, Chen KS, Park SS, Osterholm DE, Withers MA, Kimonis V, Summers AM, Meschino WS, Anyane-Yeboa K, Kashork CD, et al. *Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion.* Nat Genet **24**: 84-87 (2000).

Potocki L, Bi W, Treadwell-Deering D, Carvalho CMB, Eifert A, Friedman E, Glaze D, Krull K, Lee JA, Lewis RA, et al. *Characterization of the Potocki-Lupski syndrome [dup(17)(p11.2p11.2)] and delineation of a dosage-sensitive critical interval that can convey an autism phenotype.* Am J Hum Genet **80**: 663-649 (2007).

Quackenbush J. *Microarray data normalization and transformation.* Nat Genet **32** Suppl: S496-501 (2002).

Rabiner LR. *A tutorial on hidden Markov models and selected applications in speech recognition.* Proc IEEE Inst Electr Electron Eng **77**: 257–286 (1989).

Rademakers R, Cruts M, van Broeckhoven C. *The role of tau (MAPT) in frontotemporal dementia and related tauopathies.* Hum Mutat **24**: 277-295 (2004).

Rauch A, Ruschendorf F, Huang J, Trautmann U, Becker C, Thiel C, Jones KW, Reis A, Nurnberg P. *Molecular karyotyping using an SNP array for genomewide genotyping.* J Med Genet **41**: 916–922 (2004).

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. *Global variation in copy number in the human genome.* Nature **444**: 444-454 (2006a).

Redon R, Baujat G, Sanlaville D, Le Merrer M, Vekemans M, Munnich A, Carter NP, Cormier-Daire V, Colleaux L. *Interstitial 9q22.3 microdeletion: clinical and molecular*

characterisation of a newly recognised overgrowth syndrome. Eur J Hum Genet **14**: 759-767 (2006b).

Reiter LT, Murakami T, Koeuth T, Pentao L, Muzny DM, Gibbs RA, Lupski JR. *A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element.* Nat Genet **12**: 288-297 (1996).

Riccardi VM, Sujansky E, Smith AC, Francke U. *Chromosomal imbalance in the Aniridia-Wilms' tumor association: 11p interstitial deletion.* Pediatrics **61:** 604-610 (1978).

Rickard S, Parker M, van't Hoff W, Barnicoat A, Russell-Eggitt I, Winter RM, Bitner-Glindzicz M. *Oto-facio-cervical (OFC) syndrome is a contiguous gene deletion syndrome involving EYA1:molecular analysis confirms allelism with BOR syndrome and further narrows the Duane syndrome critical region to 1 cM.* Hum Genet **108**: 398–403 (2001).

Robin NH, Sellinger B, McDonald-McGinn D, Zackai EH, Emanuel BS, Driscoll DA. *Classical Noonan syndrome is not associated with deletions of 22q11.* Am J Med Genet **56**: 94–96 (1995).

Robinson SW, Morris CD, Goldmuntz E, Reller MD, Jones MA, Steiner RD, Maslen CL. *Missense mutations in CRELD1 are associated with cardiac atrioventricular septal defects.* Am J Hum Genet **72**: 1047–1052 (2003).

Roessler E, Belloni E, Gaudenz K, Jay P, Berta P, Scherer SW, Tsui LC, Muenke M. *Mutations in the human Sonic Hedgehog gene cause holoprosencephaly.* Nat Genet **14**: 357-360 (1996).

Romeo G, Ronchetto P, Luo Y, Barone V, Seri M, Ceccherini I, Pasini B, Bocciardi R, Lerone M, Kaariainen H, et al. *Point mutations affecting the tyrosine kinase domain of the RET proto-oncogene in Hirschsprung's disease.* Nature **367**: 377–378 (1994).

Rooms L, Reyniers E, Wuyts W, Storm K, van Luijk R, Scheers S, Wauters J, van den Ende J, Biervliet M, Eyskens F, et al. *Multiplex ligation-dependent probe amplification to detect subtelomeric rearrangements in routine diagnostics.* Clin Genet **69**: 58-64 (2006).

Rooney DE. Editor of: *Human cytogenetics: constitutional analysis.* New York: Oxford Univ Press, page 99– 127 (2001).

Rosenberg C, Knijnenburg J, Bakker E, Vianna-Morgante AM, Sloos W, Otto PA, Kriek M, Hansson K, Krepischi-Santos AC, Fiegler H, et al. *Array-CGH detection of micro rearrangements in mentally retarded individuals: clinical significance of imbalances present in both affected children and normal parents.* J Med Genet **43**: 180-186 (2006).

Sahoo T, Cheung SW, Ward P, Darilek S, Patel A, del Gaudio D, Kang SH, Lalani SR, Li J, McAdoo S, et al. *Prenatal diagnosis of chromosomal abnormalities using array-based comparative genomic hybridization.* Genet Med **8**: 719-727 (2006).

Sanlaville D, Romana SP, Lapierre JM, Amiel J, Genevieve D, Ozilou C, Le Lorch M, Brisset S, Gosset P, Baumann C, et al. *CGH study of 27 patients with CHARGE association.* Clin Genet **61**: 135-138 (2002).

Sanlaville D, Etchevers HC, Gonzales M, Martinovic J, Clement-Ziza M, Delezoide AL, Aubry MC, Pelet A, Chemouny S, Cruaud C, et al. *Phenotypic spectrum of CHARGE syndrome in fetuses with CHD7 truncating mutations correlates with expression during human development.* J Med Genet **43**: 211-217 (2006).

Schaeffer AJ, Chung J, Heretis K, Wong A, Ledbetter DH, Lese Martin C. C*omparative genomic hybridization-array analysis enhances the detection of aneuploidies and*

submicroscopic imbalances in spontaneous miscarriages. Am J Hum Genet **74**: 1168-1174 (2004).

Schmickel RD. *Contiguous gene syndromes: a component of recognizable syndromes.* J Pediatr **109**: 231-241 (1986).

Schoumans J, Anderlid BM, Blennow E, Teh BT, Nordenskjold M. *The performance of CGH array for the detection of cryptic constitutional chromosome imbalances.* J Med Genet **41**: 198-202 (2004).

Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, Nordenskjold M. Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). J Med Genet **42**: 699-705 (2005).

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. *Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.* Nucleic Acids Res **30**: e57 (2002).

Schubbert S, Zenker M, Rowe SL, Boll S, Klein C, Bollag G, van der Burgt I, Musante L, Kalscheuer V, Wehner LE, et al. *Germline KRAS mutations cause Noonan syndrome.* Nat Genet **38**: 331-336 (2006).

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, et al. *Large-scale copy number polymorphism in the human genome.* Science **305**: 525-528 (2004).

Shaikh TH, Kurahashi H, Emanuel BS. *Evolutionarily conserved low copy repeats (LCRs) in 22q11 mediate deletions, duplications, translocations, and genomic instability: an update and literature review.* Genet Med **3**: 6-13 (2001).

Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, et al. *Segmental duplications and copy-number variation in the human genome.* Am J Hum Genet **77**: 78-88 (2005).

Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC, et al. *Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome.* Nat Genet **38**: 1038-1042 (2006).

Sharp AJ, Selzer RR, Veltman JA, Gimelli S, Gimelli G, Striano P, Coppola A, Regan R, Price SM, Knoers NV, et al. *Characterization of a recurrent 15q24 microdeletion syndrome.* Hum Mol Genet **16**: 567-572 (2007).

Shaw CJ, Bi W, Lupski JR. *Genetic proof of unequal meiotic crossovers in reciprocal deletion and duplication of 17p11.2.* Am J Hum Genet **71**: 1072-1281 (2002).

Shaw CJ, Lupski JR. *Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease.* Hum Mol Genet **13**: R57-64 (2004).

Shaw CJ, Withers MA, Lupski JR. *Uncommon Smith- Magenis syndrome deletions can be recurrent by utilizing alternate LCRs as homologous recombination substrates.* Am J Hum Genet **75**: 75–81 (2004a).

Shaw CJ, Shaw CA, Yu W, Stankiewicz P, White LD, Beaudet AL, Lupski JR. *Comparative genomic hybridisation using a proximal 17p BAC/PAC array detects rearrangements responsible for four genomic disorders.* J Med Genet **41**: 113-119 (2004b).

Shaw CJ, Lupski JR. *Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms.* Hum Genet **116**:1-7 (2005).

Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H, Firth H, Sanlaville D, Winter R, Colleaux L, et al. *Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in*

patients with learning disability/mental retardation and dysmorphic features. J Med Genet **41**: 241-248 (2004).

Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, Cumming S, Dunn C, Kalaitzopoulos D, et al. *Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability.* Nat Genet **38**: 1032-1037 (2006).

Shotelersuk V, Larson D, Anikster Y, McDowell G, Lemons R, Bernardini I, Guo J, Thoene J, Gahl WA. *CTNS mutations in an American-based population of cystinosis patients.* Am J Hum Genet **63**: 1352-1362 (1998).

Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH. *Mutations in RAI1 associated with Smith-Magenis syndrome.* Nat Genet **33**: 466-468 (2003).

Slater HR, Bailey DK, Ren H, Cao M, Bell K, Nasioulas S, Henke R, Choo KH, Kennedy GC. *High resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs.* Am J Hum Genet **77**: 709-726 (2005).

Smeets DF. *Historical prospective of human cytogenetics: from microscope to microarray.* Clin Biochem **37**: 439-446 (2004).

Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, et al. *Assembly of microarrays for genome-wide measurements of DNA copy number.* Nat Genet **29**: 263-164 (2001).

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P. *Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances.* Genes Chromosomes Cancer **20**: 399-407 (1997).

Solomon NM, Ross SA, Morgan T, Belsky JL, Hol FA, Karnes PS, Hopwood NJ, Myers SE, Tan AS, Warne GL, et al. *Array comparative genomic hybridisation analysis of boys with X linked hypopituitarism identifies a 3.9 Mb duplicated critical region at Xq27 containing SOX3.* J Med Genet **41**: 669-678 (2004).

Spiteri E, Babcock M, Kashork CD, Wakui K, Gogineni S, Lewis DA, Williams KM, Minoshima S, Sasaki T, Shimizu N, et al. *Frequent translocations occur between low copy repeats on chromosome 22q11.2 (LCR22s) and telomeric bands of partner chromosomes.* Hum Mol Genet **12**: 1823-1837 (2003).

Stankiewicz P, Parka SS, Holder SE, Waters CS, Palmer RW, Berend SA, Shaffer LG, Potocki L, Lupski JR. *Trisomy 17p10-p12 resulting from a supernumerary marker chromosome derived from chromosome 17: molecular analysis and delineation of the phenotype.* Clin Genet **60**: 336-344 (2001a).

Stankiewicz P, Park S-S, Inoue K, Lupski JR. *The evolutionary chromosome translocation 4;19 in Gorilla gorilla is associated with microduplication of the chromosome fragment syntenic to sequences surrounding the human proximal CMT1A-REP.* Genome Res **1**: 1205–1210 (2001b).

Stankiewicz P, Lupski JR. *Genome architecture, rearrangements and genomic disorders.* Trends Genet **18**: 74-82 (2002).

Stankiewicz P, Shaw CJ, Dapper JD, Wakui K, ShaVer LG, Withers M, Elizondo L, Park S-S, Lupski JR. *Genome architecture catalyzes nonrecurrent chromosomal rearrangements.* Am J Hum Genet **72**: 1101–1116 (2003).

Stankiewicz P, Shaw CJ, Withers M, Inoue K, Lupski JR. *Serial segmental duplications during primate evolution result in complex human genome architecture.* Genome Res **14**: 2209–2220 (2004).

Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, et al. A common inversion under selection in Europeans. Nat Genet **37**: 129-37 (2005).

Stoeckert CJ Jr, Causton HC, Ball CA. *Microarray databases: standards and ontologies.* Nat Genet **32** Suppl: 469-473 (2002).

Takei Y, Teng J, Harada A, Hirokawa N. *Defects in axonal elongation and neuronal migration in mice with disrupted tau and map1b genes.* J Cell Biol **150**: 989-1000 (2000).

Tanabe H Muller S, Neusser M, von Hase J, Calcagno E, Cremer M, Solovei I, Cremer C, Cremer T. *Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates.* Proc Natl Acad Sci USA **99**: 4424-4429 (2002).

Tartaglia M, Mehler EL, Goldberg R, Zampino G, Brunner HG, Kremer H, van de .Burgt I, Crosby AH, Ion A, Jeffery S, et al. *Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome.* Nat Genet **29**: 465–468 (2001).

Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A. *Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer.* Genomics **13**: 718-725 (1992).

Tellier AL, Amiel J, Delezoide AL, Audollent S, Auge J, Esnault D, Encha-Razavi F, Munnich A, Lyonnet S, Vekemans M, et al. *Expression of the PAX2 gene in human embryos and exclusion in the CHARGE syndrome.* Am J Med Genet **93**: 85-88 (2000).

Temple IK, Gardner RJ, Robinson DO, Kibirige MS, Ferguson AW, Baum JD, Barber JC, James RS, Shield JP. *Further evidence for an imprinted gene for neonatal diabetes localised to chromosome 6q22-q23.* Hum Mol Genet **5**: 1117-1121 (1996).

Tjio HL, Levan A. *The chromosome numbers of man.* Hereditas **42**: 1–6 (1956).

Ton CC, Hirvonen H, Miwa H, Weil MM, Monaghan P, Jordan T, van Heyningen V, Hastie ND, Meijers-Heijboer H, Drechsler M. *Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region*. Cell **67**: 1059–1074 (1991).

Toyo-Oka K, Shionoya A, Gambello MJ, Cardoso C, Leventer R, Ward HL, Ayala R, Tsai LH, Dobyns W, Ledbetter D, Hirotsune S, Wynshaw-Boris A. *14-3-3epsilon is important for neuronal migration by binding to NUDEL: a molecular explanation for Miller-Dieker syndrome.* Nat Genet **34**: 274-285 (2003).

Trask BJ. *Human cytogenetics: 46 chromosomes, 46 years and counting.* Nat Rev Genet **3**: 769–778 (2002).

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, et al. *Fine-scale structural variation of the human genome.* Nat Genet **37**: 727-732 (2005).

Tyson C, Harvard C, Locker R, Friedman JM, Langlois S, Lewis ME, Van Allen M, Somerville M, Arbour L, Clarke L, et al. *Submicroscopic deletions and duplications in individuals with intellectual disability detected by array CGH.* Am J Med Genet A **139**: 173-185 (2005).

Valentijn LJ, Baas F, Wolterman RA, Hoogendijk JE, van den Bosch NH, Zorn I, Gabreels-Festen AW, de Visser M, Bolhuis PA. *Identical point mutations of PMP-22 in Trembler-J mouse and Charcot-Marie-Tooth disease type 1A.* Nat Genet **2**: 288-291 (1992).

Valentine H, Sergovich F. *A syndrome associated with interstitial deletion of chromosome 7q.* Birth defects **3b**: 261–262 (1977).

van Bokhoven H, Celli J, van Reeuwijk J, Rinne T, Glaudemans B, van Beusekom E, Rieu P, Newbury-Ecob RA, Chiang C, Brunner HG. *MYCN haploinsufficiency is associated with reduced brain size and intestinal atresias in Feingold syndrome.* Nat Genet **37**: 465–467 (2005).

van Buggenhout G, Melotte C, Dutta B, Froyen G, van Hummelen P, Marynen,P, Matthijs G, de Ravel T, Devriendt K, Fryns JP, et al. *Mild Wolf-Hirschhorn syndrome: micro-array CGH analysis of atypical 4p16.3 deletions enables refinement of the genotype-phenotype map.* J Med Genet **41**: 691-698 (2004).

van Deutekom JC, Wijmenga C, van Tienhoven EA, Gruter A-M, Hewitt E, Padberg GW, van Ommen G-J, Hofker MH, Frants RR. *FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit.* Hum Mol Genet **2**: 2037-2042 (1993).

van Prooijen-Knegt AC, Van Hoek J, Bauman JG, Van Duijn P, Wool IG, Van der Ploeg M. *In situ hybridization of DNA sequences inhuman metaphase chromosomes visualized by an indirect fluorescent immunocytochemical procedure.* Exp Cell Res **141**: 397–407 (1982).

Varela MC, Krepischi-Santos AC, Paz JA, Knijnenburg J, Szuhai K, Rosenberg C, Koiffmann CP. *A 17q21.31 microdeletion encompassing the MAPT gene in a mentally impaired patient.* Cytogenet Genome Res **114**: 89-92 (2006).

Veltman JA, Schoenmakers EF, Eussen BH, Janssen I, Merkx G, van Cleef B, van Ravenswaaij CM, Brunner HG, Smeets D, Geurts van Kessel A. *High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization.* Am J Hum Genet **70**: 1269-1276 (2002).

Veltman JA, Fridlyand J, Pejavar S, Olshen AB, Korkola JE, DeVries S, Carroll P, Kuo WL, Pinkel D, Albertson D, et al. *Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors.* Cancer Res **63**: 2872–2880 (2003a)

Veltman JA, Jonkers Y, Nuijten I, Janssen I, Van Der Vliet W, Huys E, Vermeesch J, Van Buggenhout G, Fryns JP, Admiraal R, et al. *Definition of a critical region on chromosome 18 for congenital aural atresia by array CGH.* Am J Hum Genet **72**: 1578–1584 (2003b).

Veltman JA, Yntema HG, Lugtenberg D, Arts H, Briault S, Huys EH, Osoegawa K, de Jong P, Brunner HG, Geurts van Kessel A, et al. *High resolution profiling of X chromosomal aberrations by array comparative genomic hybridisation.* J Med Genet **41**: 425-432 (2004).

Veltman JA. *Genomic Microarrays in Clinical diagnosis.* Curr Op Ped **18**: 598-603 (2006).

Vermeulen S, Menten B, Van Roy N, Van Limbergen H, De Paepe A, Mortier G, Speleman F. *Molecular cytogenetic analysis of complex chromosomal rearrangements in patients with mental retardation and congenital malformations: delineation of 7q21.11 breakpoints.* Am J Med Genet A **124**: 10-18 (2004).

Visser R, Shimokawa O, Harada N, Kinoshita A, Ohta T, Niikawa N, Matsumoto N. *Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion.* Am J Hum Genet **76**: 52-67 (2005).

Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, Straatman H, van der Vliet W, Huys EH, van Rijk A, et al. *Array-based comparative genomic*

hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet **73**: 1261-1270 (2003).

Wallis DE, Roessler E, Hehr U, Nanni L, Wiltshire T, Richieri-Costa A, Gillessen-Kaesbach G, Zackai EH, Rommens J, Muenke M. *Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly.* Nat Genet **22**: 196-198 (1999).

Wang NJ, Liu D, Parokonny AS, Schanen NC. *High-resolution molecular characterization of 15q11-q13 rearrangements by array comparative genomic hybridization (array CGH) with detection of gene dosage.* Am J Hum Genet **75**: 267-281 (2004).

Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G. *Human diallelic insertion/deletion polymorphisms.* Am J Hum Genet **71**: 854-862 (2002).

Weise A, Rittinger O, Starke H, Ziegler M, Claussen U, Liehr T. *De novo 9-break-event in one chromosome 21 combined with a microdeletion in 21q22.11 in a mentally retarded boy with short stature.* Cytogenet Genome Res **103**: 14-16 (2003).

Wessendorf S, Schwaenen C, Kohlhammer H, Kienle D, Wrobel G, Barth TF, Nessling M, Moller P, Dohner H, Lichter P, et al. *Hidden gene amplifications in aggressive B-cell non-Hodgkin lymphomas detected by microarraybased comparative genomic hybridization.* Oncogene **22**: 1425–1429 (2003).

White SJ, Breuning MH, Den Dunnen JT. *Detecting copy number changes in genomic DNA: MAPH and MLPA.* Methods Cell Biol **75**: 751-768 (2004).

Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, Gruter A-M, Hofker MH, Moerer P, Williamson R, et al. *Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy.* Nat Genet **2**: 26-30 (1992).

Wilhelm M, Veltman JA, Olshen AB, Jain AN, Moore DH, Presti JC, Kovacs G, Waldman FM. *Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer.* Cancer Res **62**: 957-960 (2002).

Willatt L, Cox J, Barber J, Cabanas ED, Collins A, Donnai D, FitzPatrick DR, Maher E, Martin H, Parnau J, et al. *3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome*. Am J Hum Genet **77**: 154-160 (2005).

Wirtenberger M, Hemminki K, Burwinkel B. *Identification of frequent chromosome copy-number polymorphisms by use of high-resolution single-nucleotide-polymorphism arrays.* Am J Hum Genet **78**: 520-522 (2006).

Woodage T, Basrai MA, Baxevanis AD, Hieter P, Collins FS. *Characterization of the CHD family of proteins.* Proc Natl Acad Sci USA **94**: 11472-11477 (1997).

Woods KS, Cundall M, Turton J, Rizotti K, Mehta A, Palmer R, Wong J, Chong WK, Al Zyoud M, El Ali M, et al. *Over- and underdosage of SOX3 is associated with infundibular hypoplasia and hypopituitarism.* Am J Hum Genet **76**: 833-849 (2005).

Woodward K, Malcolm S. *Proteolipid protein gene: Pelizaeus-Merzbacher disease in humans and neurodegeneration in mice.* Trends Genet **15**: 125-128 (1999).

Wu YQ, Nickerson E, Shaffer LG, Keppler-Noreuil K, Muilenburg A. *A case of Williams syndrome with a large, visible cytogenetic deletion.* J Med Genet **36**: 928–932 (1999).

Yatsenko SA, Treadwell-Deering D, Krull K, Glaze D, Horz M, Stankiewicz P, Lupski JR, Potocki L. *Trisomy 17p10-p12 due to mosaic supernumerary marker chromosome: delineation of molecular breakpoints and clinical phenotype and comparison to other proximal 17p segmental duplications.* Am J Med Genet **138A**: 175–180 (2005).

Yu W, Ballif BC, Kashork CD, Heilstedt HA, Howard LA, Cai WW, White LD, Liu W, Beaudet AL, Bejjani BA, et al. *Development of a comparative genomic hybridization microarray*

and demonstration of its utility with 25 well-characterized 1p36 deletions. Hum Mol Genet **12**: 2145-2152 (2003).

Zafarana G, Grygalewicz B, Gillis AJM, Vissers LE, van der Vliet W, van Gurp RJ, Debiec-Rychter M, Oosterhuis JW, Geurts van Kessel A, Schoenmakers EF, et al. *12p-amplicon structure analysis in testicular germ cell tumors of adolescents and adults by array CGH.* Oncogene **22**: 7695–7701 (2003).

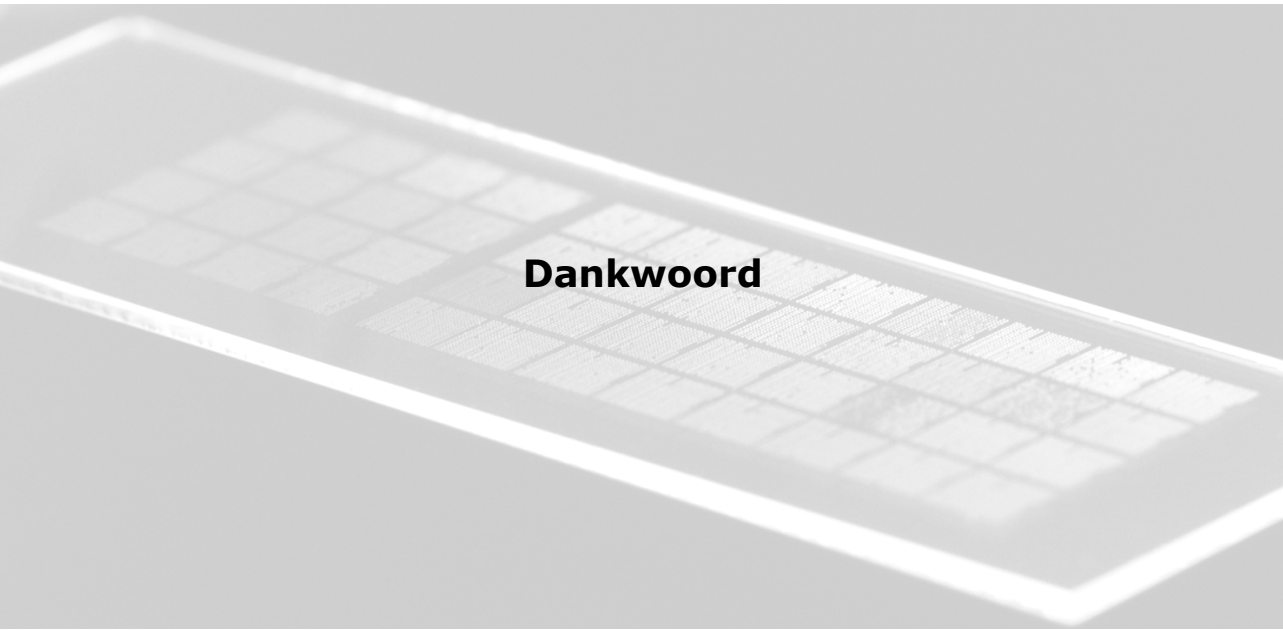
Zhang J, Feuk L, Duggan GE, Khaja R, Scherer SW. *Development of bioinformatic resources for display and analysis of copy number and other structural variants in the human genome.* Cytogenet Genome Res **115**: 205-215 (2006).

Zhou J, Mochizuki T, Smeets H, Antignac C, Laurila P, De Paepe A, Tryggvason K, Reeders ST. *Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in inherited smooth muscle tumors.* Science **261**: 1167–1169 (1993).

Zody MC, Garber M, Sharpe T, Young SK, Rowen L, O'Neill K, Whittaker CA, Kamal M, Chang JL, Cuomo CA, et al. *Analysis of the DNA sequence and duplication history of human chromosome 15.* Nature **440**: 671-675 (2006).

URLs for data presented in this thesis are as follows:

Affymetrix
http://www.affymetrix.com
BACPAC Resources Center
http://www.chori.org/bacpac/
BACPAC Resources Center's Human BAC Minimal Tiling Set Web site:
http://bacpac.chori.org/pHumanMinSet.htm
Database of Genomic Variants
http://projects.tcag.ca/variation/
DECIPHER
http://www.sanger.ac.uk/PostGenomics/decipher/
ECARUCA
http://www.ecaruca.net/
Ensembl
http://www.ensembl.org
Human Genome Variation database
http://hgvbase.cgb.ki.se/
MRC-Holland
http://www.mlpa.com/index.htm
NCBI Blast2
http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html
Online Mendelian Inheritance in Man (OMIM)
http://www.ncbi.nlm.nih.gov/Omim/
Repeatmasker through UCSC genome browser:
http://genome.ucsc.edu



Dankwoord

Het dankwoord schrijven van mijn proefschrift is het laatste wat ik moet doen voordat het naar de drukker kan. Dit is het moment waarop je nog meer dan anders beseft dat dit proefschrift veel meer is dan alleen mijn promotieonderzoek, het beschrijft een enorme groepsprestatie. Zonder een microarray was dit boekje er nooit geweest, maar er zijn maar weinig mensen die weten hoeveel tijd en mankracht het kost om een microarray te maken… laat staan om daarna nog eens uit te zoeken wat nu het beste protocol is om er mee te werken, welke patiënten het meest geschikt zijn om te onderzoeken, en welke software ontwikkeld moet worden om de immens grote datasets te analyseren. Er zijn dan ook zo veel mensen betrokken bij deze groepsprestatie, dat ik wil beginnen om iedereen te bedanken, zo weet ik ook meteen dat ik niemand vergeet. Iedereen… BEDANKT!

Uiteraard zijn er mensen geweest die op een speciale manier hebben bijgedragen aan dit onderzoek, en hen wil ik graag persoonlijker bedanken.

Beste Ad, als mijn promotor sta jij boven aan dit lijstje. Aan het einde van mijn afstudeerstage was jij degene die me al wandelend op de gang vroeg of ik niet wilde blijven om promotieonderzoek te komen doen. Natuurlijk hoefde ik daar niet lang over na te denken. Bedankt voor het steeds opnieuw lezen en verbeteren van weer een nieuwe versie van een manuscript. Ik durfde je in de laatste fase op vrijdagmiddagen de stukken niet eens meer te e-mailen omdat ik wist dat je er dan in het weekend mee aan de slag zou gaan zodat ik op maandag meteen weer verder kon. En zie hier, het boekje is af!

Beste Joris, trotse copromotor van je eerste AIO, zonder jouw aanstekelijke enthousiasme voor het onderzoek, had ik het vast een stuk minder leuk gevonden. Ik ben dan ook trots dat jij mijn begeleider bent geweest. Door jouw pragmatische manier van denken en handelen heeft te allen tijde de vaart in het onderzoek gezeten, en bleef ik niet te lang hangen in ‘zinloze’ details. Inmiddels weet ik dat ‘goed’ vaak ‘goed genoeg’ is, en dat ‘perfect’ iets is dat niet bestaat. Bedankt voor al je vertrouwen in mij, vooral bij de voorbereidingen van de presentatie voor de ASHG. Tijdens de voorbereidingen van mijn eerste literatuurbespreking moest je letterlijk de deur op slot draaien omdat ik het verschrikkelijk eng vond. Wie had dan kunnen voorzien dat het geven van een presentatie het hoogtepunt van mijn promotieonderzoek zou worden?! Zonder jouw steun was dit nooit gelukt.

Beste Han, bedankt voor de kans om binnen de afdeling Antropogenetica dit promotieonderzoek te mogen doen, en voor al je goede ideeën en adviezen die je me tijdens het onderzoek hebt gegeven.

Beste collega’s van de microarray groep, Irene, Walter, Simon, Roland, Rolph, Eveline, Marloes, Terry, Suzanne, Ramprasath, Eric, Ine, Ineke, Diederick, Jayne, Christian, Michael en Markus, maar ook iedereen die er nu niet meer werkt, Judith, Gerard, Yvonne, Martin, Martijn, Huub, Ton en Corine, jullie wil ik bedanken voor alles wat jullie gedaan hebben voor dit proefschrift. Jullie verdienen een speciaal plekje want jullie waren degene die al die BAC clones voor de 32K set bij elkaar pipetteerden, de arrays print(t)en, zorg(d)en voor het optimale protocol en validatie van alle afwijkingen. Als iemand zich ooit afvraagt wat de betekenis is van de uitdrukking “There is no I in TEAM” moeten ze eens naar deze groep komen kijken. Ik ben blij om in een dergelijk team onderzoek te mogen

doen. Een niet onbelangrijk deel van ons TEAM vind ik de gezelligheid die altijd aanwezig is. Vooral de koffie 's ochtends met iedereen is een goed begin van de dag. Ik wil ook iedereen bij tumorcytogenetica, Lilian, Lianne, Klaas, Linda, Diederik, Anke, en degene die intussen naar elders zijn gegaan, Imke, Helma, Jan, José, Anita, Erik, Marga, Anke, Marieke, Daniëlle, Marc, en Monique, bedanken voor alle hulp, met name bij het maken van de eerste generatie microarrays.

Beste Conny, Bert, David, Ilse, en andere klinisch genetici binnen de afdeling, bedankt voor jullie inzet om de juiste patiënten bij elkaar te verzamelen. Zonder jullie uitgebreide patiëntenselectie was het onmogelijk geweest om de array CGH technologie in te zetten binnen het vakgebied van de klinische genetica. Deze gedetailleerde klinische analyse heeft ervoor gezorgd dat zowel een bekend syndroom kon worden opgelost, als een nieuw syndroom kon worden gevonden. In het kader van het eerste project wil ik ook graag dr. Admiraal bedanken voor het in kaart brengen van de gehoorsafwijkingen bij de patiënten met CHARGE syndroom. I thank dr. Hurst for providing DNA of the t(6;8)(6p8p;6q8q) patient.

Natuurlijk ook iedereen van de celkweek bedankt voor het telkens weer opgroeien van alle cellijnen en het aanleveren van de celpellets. Vanzelfsprekend ook iedereen bij cyto-diagnostiek bedankt voor het verder verwerken van de celpellets tot metafase-preparaten.

Dear Paweł and Jim, a special thanks to you for having me in Houston. When I started my PhD project, I read all your papers and was overwhelmed by your impressive achievements and contributions to science. I would never have dreamt of actually working with you. Or come to that, publish a joint paper that joins us forever in PubMed! I would like to thank you for what you taught me, and for making my stay in Houston a great one. Also, thank you Beata and Marta for teaching me the true meaning of "shop-till-you-drop"!

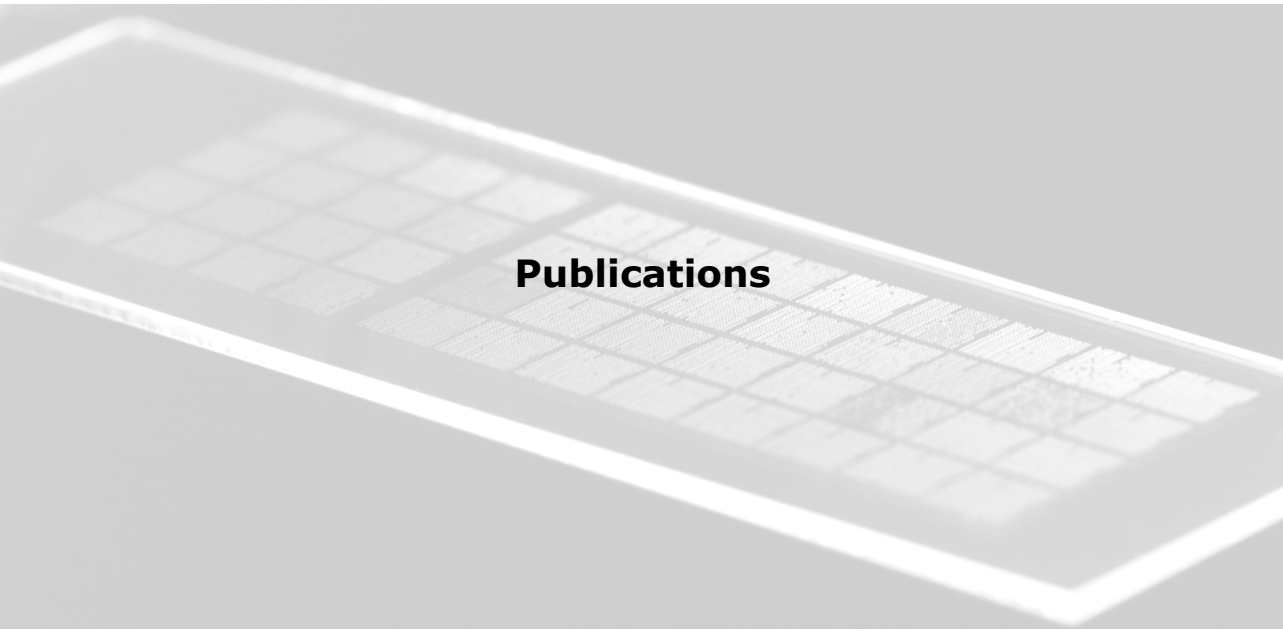
Ook wil ik dr. Stefan White, dr. den Dunnen, prof. dr. Breuning uit Leiden bedanken voor de prettige samenwerking. Beste Stefan, het heeft heel wat voeten in aarde gehad, maar het is ons uiteindelijk toch nog gelukt om onze paper te publiceren.

Beste Mirjam en Hanka, jullie wil ik bedanken voor al het werk dat jullie me uit handen hebben genomen tijdens het uitvoeren van jullie stageprojecten. Ik heb er bewondering voor hoe jullie beiden hebben laten zien dat een carrière als topsportster en onderzoekster te combineren is. Mirjam, zet hem op volgend jaar op de Olympische Spelen. Ik zal beloven samen met Joris naar de wedstrijden te kijken en hem dan nogmaals uitleggen dat het geen volleybal*elftal* is maar een volleybal*team*. Hanka, zet hem op tijdens je promotieonderzoek. Als het jou niet lukt om het gat tussen de informatici en biologen te dichten, dan lukt het niemand. Ik weet zeker dat je er in zult slagen.

Lieve Linda en Mariëlle, ik ben ontzettend blij dat jullie mijn paranimfen willen zijn. Mariëlle het was heerlijk om buiten het werk om met iemand te kunnen kletsen over het onderzoek, wetende dat je begreep waar ik het over had. Linda, het was heerlijk om met je over het onderzoek te kletsen, wetende dat je vaak geen flauw idee had waar ik het over had. Ik heb genoten van de weekendjes skiën, Parijs, en alle etentjes en hockeytoernooien. Dat jullie hier vandaag naast mij staan is geweldig!

Lieve pap en mam, ontzettend bedankt voor jullie onvoorwaardelijke steun. Zonder deze steun was ik nooit zo ver gekomen als ik nu ben. Mam, stiekem is dit boekje ook een beetje jouw boekje omdat je altijd gezegd hebt dat, als je een beroep zou mogen kiezen, je zou doen wat ik nu doe. Pap, ontzettend bedankt voor alle keren dat jij en mama klaar stonden om me weer te verhuizen in Nijmegen. Nooit geweten dat je vijf keer kunt verhuizen binnen een halve vierkante kilometer. Maar nu is dat over… ik heb nu met Mark mijn eigen huisje gevonden. Ik ben jullie ontzettend dankbaar voor alles. Lieve Marinka en Marty, het is super om te zien dat jullie nu samen eindelijk het geluk gevonden hebben wat jullie zo verdiend hebben. Er is niets mooiers om dat geluk in december bekroond te zien worden met mijn eerste neefje of nichtje. Dear Ron, Yvonne and Darren, thank you for all your love and support. You are the best parents-in-law and brother-in-law one could wish for. I have really enjoyed our Christmas in Cornwall. I wish we lived closer so we could do such things more frequently.

Lieve Mark, jij bent de laatste in het rijtje. Je Nederlands is goed genoeg om te weten 'lest best'. Jou wil ik bedanken omdat je er altijd voor me bent. Je hebt me de rust en ruimte geven als het nodig was, ook als dat op zondagmiddag was. Nu is dat niet meer nodig, het proefschrift is af. Tijd om extra veel van elkaar te gaan genieten in ons eigen paleisje.



Publications

2003

Zafarana G, Grygalewicz B, Gillis AJ, **Vissers LELM**, van de Vliet W, van Gurp RJ, Stoop H, Debiec-Rychter M, Oosterhuis JW, Geurts van Kessel A, Schoenmakers EF, Looijenga LH, Veltman JA. 12p-amplicon structure analysis in testicular germ cell tumors of adolescents and adults by array CGH. Oncogene. 22: 7695-7701 (2003).

Vissers LELM, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, Straatman H, van der Vliet W, Huys EH, van Rijk A, Smeets D, van Ravenswaaij-Arts CM, Knoers NV, van der Burgt I, de Jong PJ, Brunner HG, Geurts van Kessel A, Schoenmakers EF, Veltman JA. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet 73: 1261-1270 (2003).

Veltman IM, Veltman JA, Arkesteijn G, Janssen IM, **Vissers LELM**, de Jong PJ, Geurts van Kessel A, Schoenmakers EF. Chromosomal breakpoint mapping by arrayCGH using flow-sorted chromosomes. Biotechniques 35: 1066-1070 (2003).

2004

Koolen DA, **Vissers LELM**, Nillesen W, Smeets D, van Ravenswaaij CM, Sistermans EA, Veltman JA, de Vries BD. A novel microdeletion, del(2)(q22.3q23.3) in a mentally retarded patient, detected by array-based comparative genomic hybridization. Clin Genet 65: 429-432 (2004).

Vissers LELM, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van der Vliet WA, Huys EH, de Jong PJ, Hamel BC, Schoenmakers EF, Brunner HG, Veltman JA, Geurts van Kessel A. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. Nat Genet 36: 955-957 (2004).

2005

Veltman JA, **Vissers LELM**. Molecular karyotyping by means of array CGH: Linking gene dosage alterations to disease phenotypes. In: DNA Microarrays (Nuber, U), BIOS Scientific Publishers), Abingdon, UK (2005).

De Vries BBA*, Pfundt R*, Leisink M, Koolen DA, **Vissers LELM**, Janssen IM, Reijmersdal S, Nillesen WM, Huys EHLPG, de Leeuw N, Smeets D, Sistermans EA, Feuth T, van Ravenswaaij-Arts CMA, Geurts van Kessel A, Schoenmakers EFPM, Brunner HB, Veltman JA. Diagnostic genome profiling in mental retardation. Am J Hum Genet 77: 606-616 (2005).

Vissers LELM, Veltman JA, Geurts van Kessel A, Brunner HB. Identification of disease genes by whole genome CGH arrays. Human Mol Genet 14: R215-223 (2005) *(Invited Review)*.

2006

Jongmans MCJ, Admiraal RJ, van der Donk KP, **Vissers LELM**, Baas AF, Kapusta L, van Hagen JM, Donnai D, de Ravel TJ, Veltman JA, Geurts van Kessel A, De Vries BBA, Brunner HG, Hoefsloot LH, van Ravenswaaij CMA. CHARGE syndrome: the phenotypic spectrum of mutations in the *CHD7* gene. J Med Genet 43: 306-314 (2006).

Koolen DA, **Vissers LELM**, Pfundt R, de Leeuw N, Knight SJL, Regan R, Kooy FR, Reyniers E, Romano C, Fichera M, Schinzel A, Baumer A, Anderlid B-M, Schoumans J, Knoers NV, Geurts van Kessel A, Sistermans EA, Veltman JA, Brunner HG, de Vries BBA. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet 38: 999-1001 (2006).

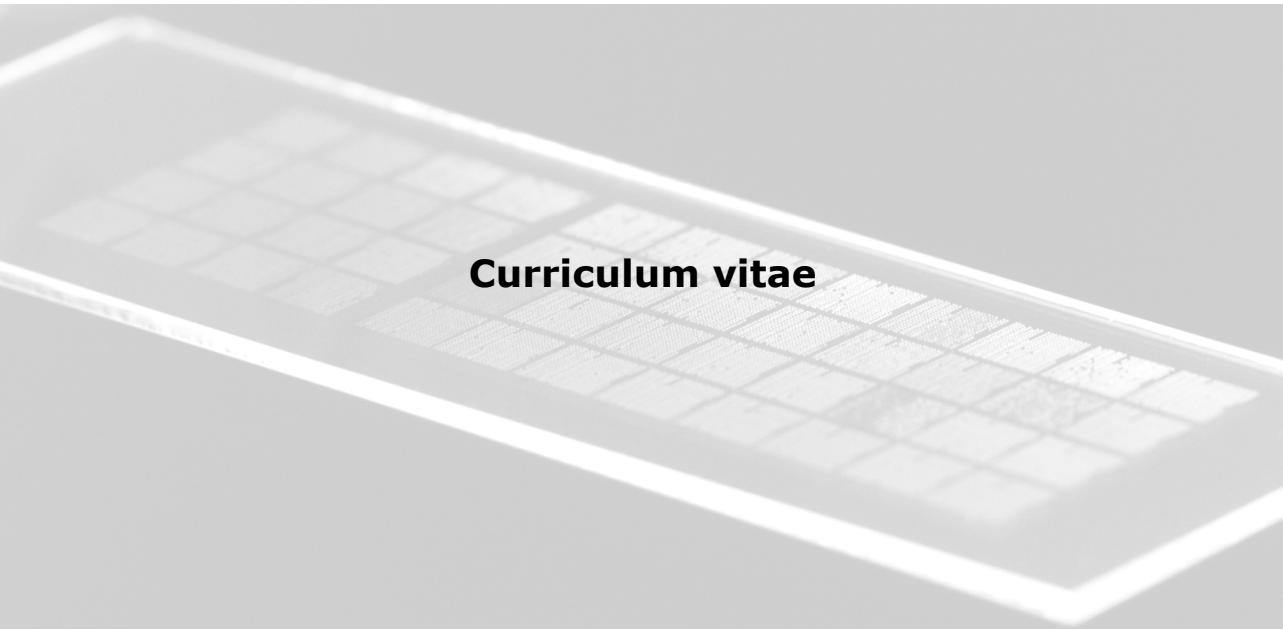
2007

Schlaubitz S, Yatsenko SA, Smith LD, Keller KL, **Vissers LELM**, Scott DA, Cai WW, Reardon W, Abdul-Rahman OA, Lammer EJ, Lifchez CA, Magenis E, Veltman JA, Stankiewicz P, Zabel BU and Lee B. Ovotestes and XY Sex Reversal in a Female with an Interstitial 9q33.3-q34.1 Deletion Encompassing NR5A1 and LMX1B causing features of Genitopatellar Syndrome. Am J Med Genet A 143: 1071-1081 (2007).

Vissers LELM, Stankiewicz P, Yatsenko SA, Crawford E, Creswick H, Gutter EM, de Vries, BBA, Pfundt R, Marcelis CLM, Zackowski J, Geurts van Kessel A, Lupski JR, Veltman JA. Complex chromosome rearrangements in 17p are associated with genomic architecture involving low-copy repeats. Hum Genet 121: 697-709 (2007).

White SJ*, **Vissers LELM***, Geurts van Kessel A, Kalay E, Lehesjoki AE, Giordano PC, van de Vosse E, Breuning MH, Brunner HG, den Dunnen JT, Veltman JA. Frequency of common copy number variation in five different ethnic populations. Cytogenet Genome Res: in press (2007).

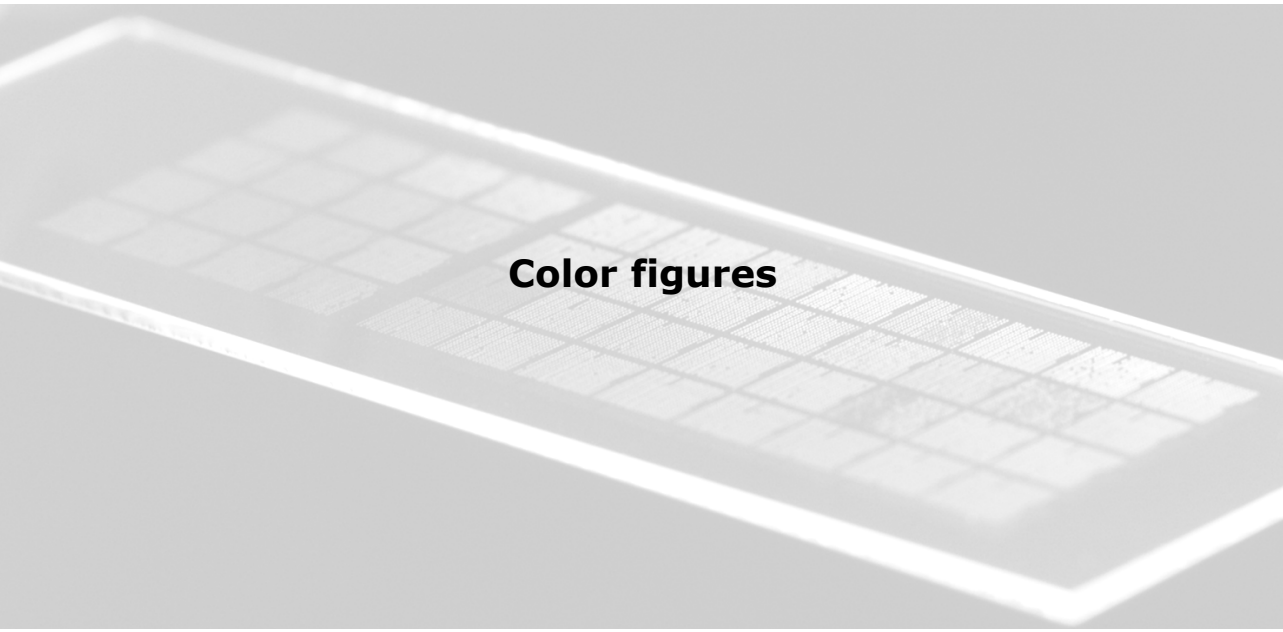
Feenstra I, **Vissers LELM**, Orsel M, Brunner HG, Geurts van Kessel A, Veltman JA, van Ravenswaaij CMA. Genotype-Phenotype mapping of chromosome 18q deletions by high-resolution array CGH; an update of the phenotypic map. Am J Med Genet A: in press (2007).



Curriculum vitae

Lisenka Elisabeth Louisa Maria Vissers werd geboren op 27 augustus 1980 te Vught, Nederland. In 1998 voltooide zij haar VWO opleiding aan het R.K. Gymnasium Beekvliet in Sint Michielsgestel. Aansluitend begon zij met de studie Biologie aan de Radboud Universiteit Nijmegen. De eerste stage binnen deze opleiding werd verricht bij de afdeling Neurologie van het UMC St. Radboud, Nijmegen. Vervolgens deed zij haar afstudeerstage bij de afdeling Antropogenetica van hetzelfde medisch centrum. In augustus 2002 werd het doctoraal examen Medische Biologie afgelegd. Van september 2002 tot september 2006 was zij aangesteld als promovenda bij de afdeling Antropogenetica van het UMC St. Radboud Nijmegen (promotor Prof. dr. A. Geurts van Kessel) onder supervisie van dr. ir. J.A. Veltman (copromotor). Tevens werd een deel van het onderzoek uitgevoerd bij de afdeling Molecular and Human Genetics van het Baylor College of Medicine te Houston, Verenigde Staten, onder directe supervisie van Prof. dr. J.R. Lupski en dr. P. Stankiewicz. Tijdens haar promotie deed zij onderzoek naar de toepasbaarheid van array CGH voor de detectie van (submicroscopische) chromosoomafwijkingen in toegepast klinisch genetisch onderzoek en fundamenteel genoomonderzoek. Tijdens haar promotieonderzoek ontving zij het Dr. I.B.M. Frye stipendium van de Radboud Universiteit Nijmegen (2005). Tevens werd aan haar een fellowship van het Netherlands Genomics Initiative toegekend (2005). De resultaten van het onderzoek zijn beschreven in dit proefschrift. Sinds januari 2007 is zij werkzaam als postdoc bij de afdeling Antropogenetica van het UMC St. Radboud, Nijmegen. Lisenka woont samen met Mark Peart.

Lisenka Elisabeth Louisa Maria Vissers was born on August 27, 1980 in Vught, the Netherlands. In 1998 she passed secondary education at the R.K. Gymnasium Beekvliet in Sint Michielsgestel. In the same year, she started to study Biology at the Radboud University Nijmegen. Her first research project was performed at the department of Neurology of the UMC St. Radboud, Nijmegen. Subsequently, she performed a research project at the department of Human Genetics at the same medical center. In August 2002, she obtained her Master of Science degree in Medical Biology. From September 2002 till September 2006, she performed a Ph.D. project at the department of Human Genetics at the UMC St. Radboud Nijmegen (promotor Prof. dr. A. Geurts van Kessel) under supervision of dr. ir. J.A. Veltman (copromotor). Part of the project was performed at the department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA, under supervision of Prof. dr. J.R. Lupski and dr. P. Stankiewicz. The aim of her Ph.D. project was to explore the feasibility of the array CGH technology for the delineation of DNA copy number alterations for applied clinical genetic and basic genome research purposes. During her Ph.D. project she was rewarded the Dr. I.B.M. Frye Stipend from the Radboud University Nijmegen (2005). In addition, she obtained a fellowship from the Netherlands Genomics Initiative (2005). The results of the project are described in this thesis. Since January 2007 she is employed as a postdoctoral fellow by the department of Human Genetics at the UMC St. Radboud, Nijmegen. Lisenka lives together with Mark Peart.



Color figures

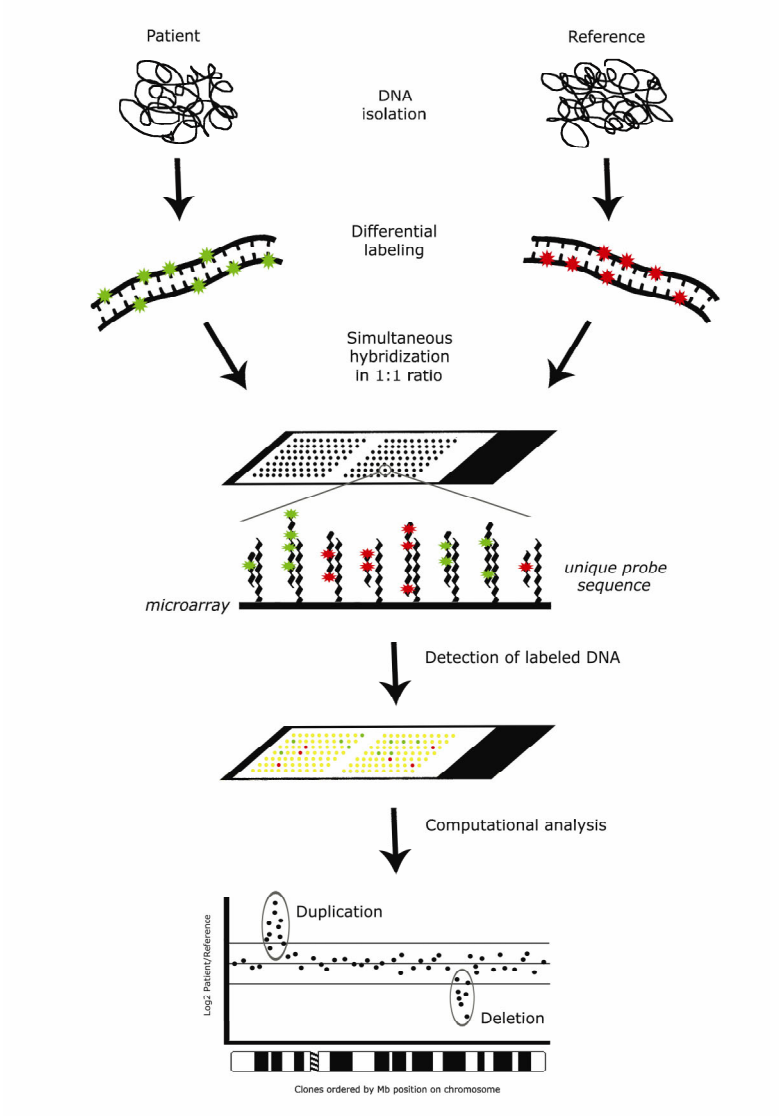


Figure 1.2
Schematic overview of the array CGH procedure. Genomic DNA samples from a test (patient; left) and reference (normal control; right) are differentially labeled with different fluorochromes, usually Cy3 and Cy5 (for green and red, respectively). The two DNA samples are mixed in equal amounts and hybridized to the microarray, onto which large-insert clone DNAs (e.g. BACs) have been robotically spotted as targets. Subsequent computer imaging assesses the relative fluorescence levels of each labeled DNA for each array target. Clones to which equal amounts of patient DNA and reference DNA have been hybridized, will appear in yellow, clones deleted in the patient DNA compared to the reference DNA will appear in red, and clones that are duplicated in the patient DNA compared to the reference DNA will appear in green. The ratio between the test and reference DNA for each clone on the array can be plotted to visualize chromosome imbalances, marked by a deviation from the normal $\log_2$ T/R ratio.

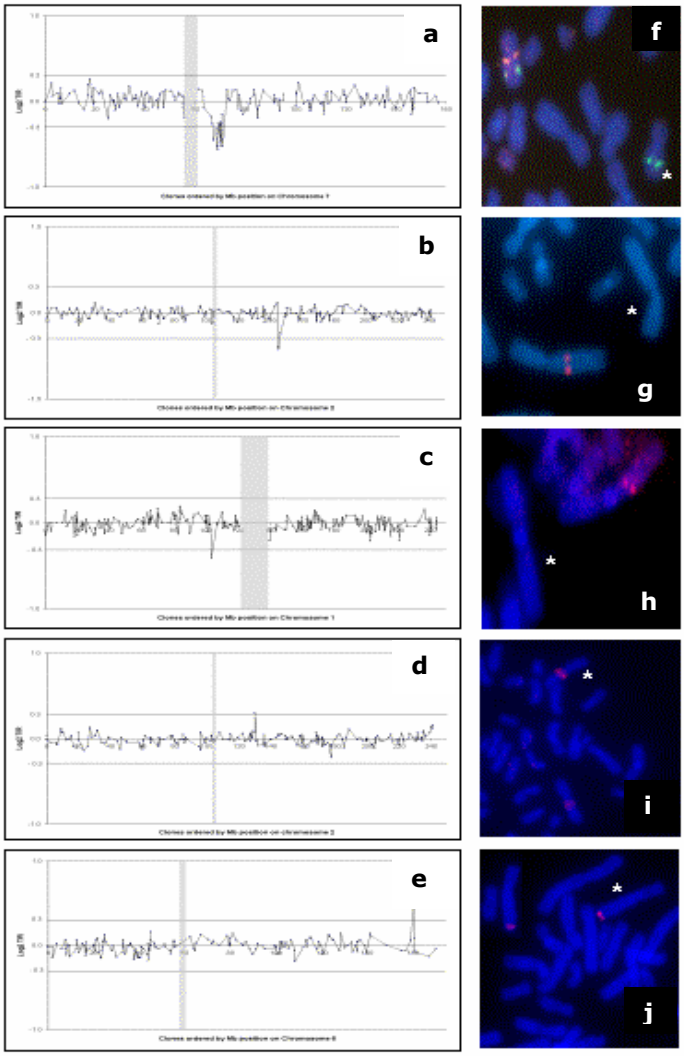


Figure 2.2
Detailed genomic profiles and FISH validation of copy number abnormalities identified in five cases with unexplained mental retardation. Panels **a–e** represent individual profiles of the affected chromosomes for each case, with clones ordered, for each chromosome, from pter to qter, on the basis of the physical mapping positions obtained from the November 2002 freeze of the UCSC genome browser. The centromeric region is indicated by a vertical gray dash, the thresholds for copy number gain (log_2T/R value 0.3) and copy number loss (log_2 T/R value -0.3) are indicated by horizontal lines. Panels **f-j** represent the FISH validation using (one of) the target clone(s) identified by array CGH. Affected chromosomes are indicated by an asterisk (*). Panel **a** shows the deletion on 7q11 in patient 1, with 14 clones in this region showing an average log_2 intensity ratio of -0.5. FISH validation of this case is shown in the adjacent panel **f**, in which one of the deleted clones on 7q11 is shown in red and an undeleted control probe is shown in green. Panel **b** shows the microdeletion on 2q22 in patient 2 with a total of three clones crossing the threshold for copy number loss, with FISH validation in the adjacent panel **g**. Deletion of a single clone on 1p21 is shown in panel **c** for patient 3; this clone was confirmed by FISH to be deleted not only in the patient (panel **h**) but also in the father of the patient. Copy number gain detected in a single clone is shown in panels **d** and **e** for patients 4 and 5, with FISH confirmation in panels **i** and **j**.

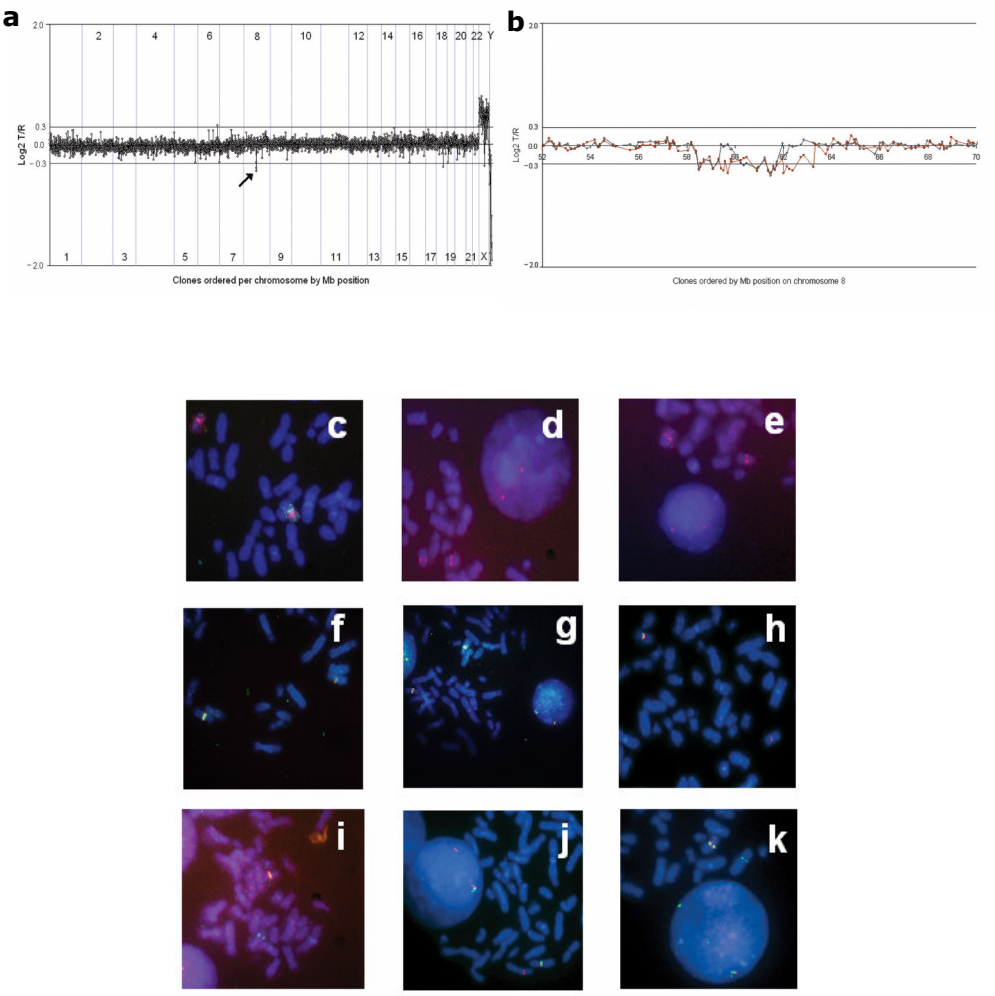


Figure 3.1
Array CGH profiles of two individuals with CHARGE syndrome and FISH validation. (**a**) Array CGH genome-wide profile of the index individual with a copy number deletion of three adjacent clones on 8q12 (arrow). This profile represents the result of a single hybridization experiment; analysis of the replicate experiment identified only the 8q12 clones as being reproducibly deleted. Vertical lines indicate chromosome boundaries. (**b**) Profile of the same individual (brown squares) on the tiling chromosome 8 BAC array with 31 clones characterizing the deletion, and the chromosome 8 profile of an individual (black circles) with an apparently balanced t(6;8) translocation that overlaps with the deletion of the index individual. (**a**, **b**) Clones are ordered on the *x* axis according to physical mapping positions; $\log_2$ transformed test-over-reference ratios for each clone are given on the *y* axis. (**c**-**g**) FISH validation of the deletion in individual 1 (**c**) and testing for *de novo* occurrence in the accompanying parents (**d-e**) with clone RP11-91I20. (**f**) Determination of proximal deletion boundary (RP11-310L18 in red is normal; RP11-44D19 in green is deleted) and (**g**) determination of distal deletion boundary; RP11-274C23 in green is deleted whereas RP11-252M13 in red is normal. (**h-k**) FISH validation of the two distinct deletions in individual 2 and determination of the deletion boundaries. (**h**) The proximal deletion contains RP11-44D19 (green) on the centromeric site but not RP11-310L18 (red), and on the telomeric site (**i**) RP11-696A12 (red) but not RP11-630K16 (green). (**j**) Normal region between the two deletions is indicated by RP11-360N16 (red), whereas RP11-91I20 (green) marks the distal deletion. (**k**) This distal deletion contains RP11-265K14 (red) on the telomeric site, but not RP11-317H6 (green).

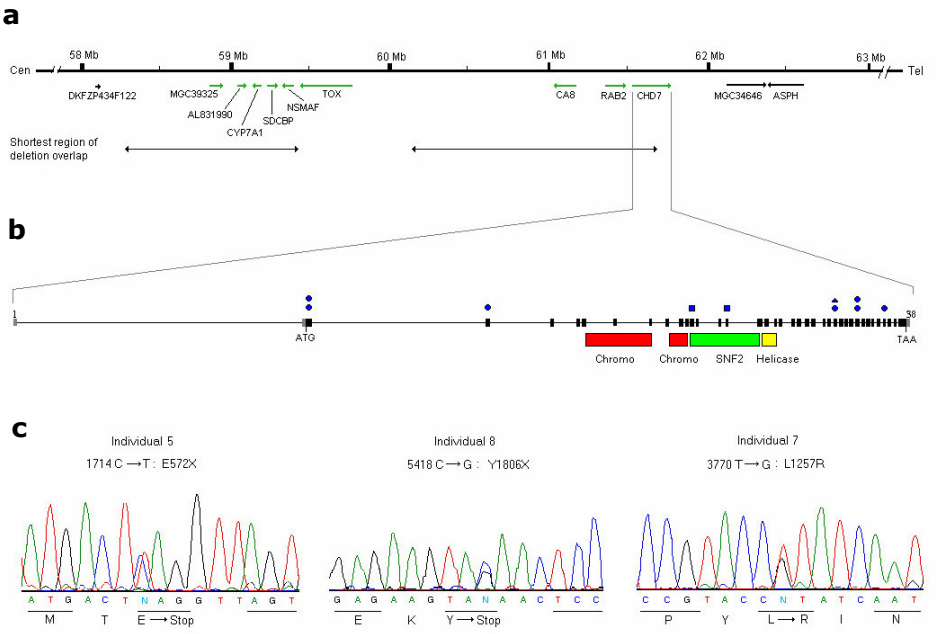


Figure 3.2
Detailed genomic view of 8q12, organization of the *CHD7* gene and mutations detected. (**a**) Transcript map of the deleted 8q12 genomic region. The shortest region of deletion overlap in the two individuals is shown. Genes in green were screened for mutations. Cen, centromeric; Tel, telomeric. (**b**) Genomic structure of *CHD7* indicating the positions of seven nonsense mutations (circles), two missense mutations (squares) and one intron-exon boundary mutation (triangle). The corresponding functional CHD7 domains are marked (colored bars). (**c**) Partial electropherograms obtained by direct sequencing of PCR products showing two nonsense mutations in individual 5 (1714C>T) and individual 8 (5418C>G) and one missense mutation in individual 7 (3770T>G).

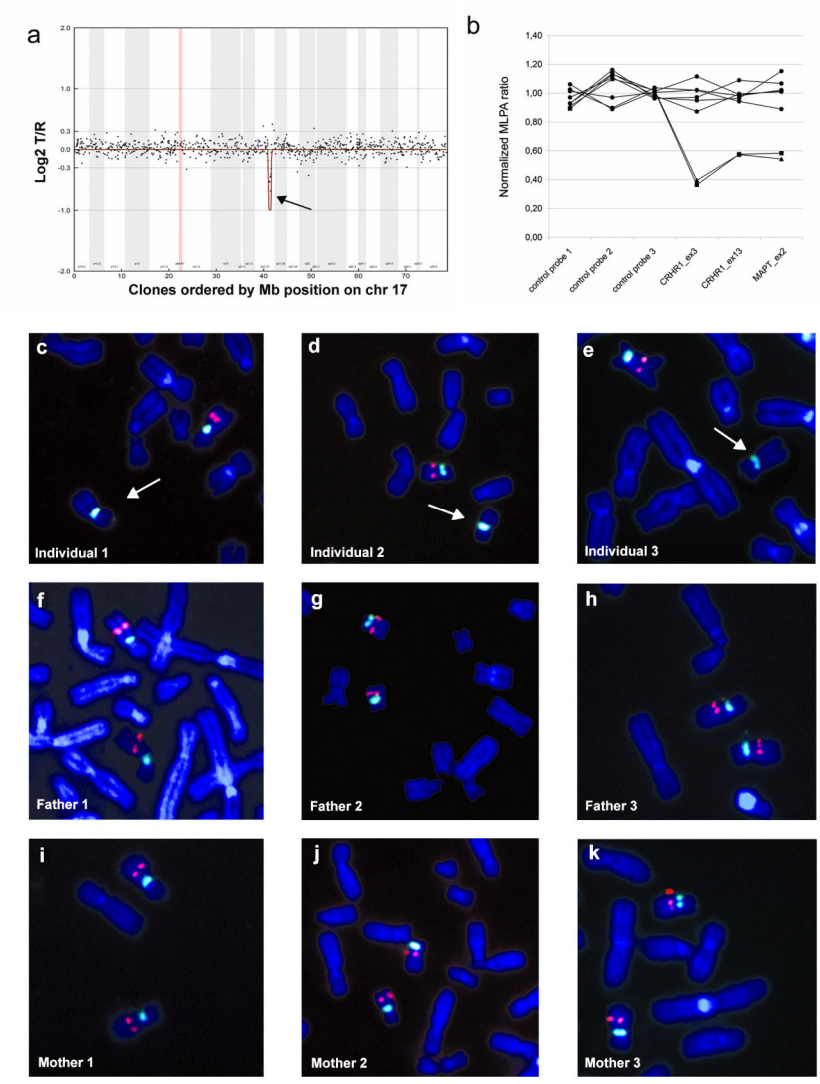


Figure 4.2
Analysis of individuals with the 17q21.31 microdeletion syndrome. (**a**) Chromosome 17 array CGH profile of individual 1 with a copy number loss of eight adjacent BAC clones on 17q21.31 (arrow). Clones are ordered on the *x* axis according to physical mapping positions; $\log_2$-transformed test-over-reference (T/R) intensity ratios for each clone are given on the *y* axis. (**b**) Multiplex ligation-dependent probe amplification (MLPA) analysis, showing a deletion of two probes in *CRHR1* and one probe in *MAPT* for both individual 2 (triangles) and individual 3 (squares). The circles represent the MLPA ratios of healthy controls. FISH validation of the 17q21.31 deletion in individuals 1, 2 and 3 (**c-e**) and testing for *de novo* occurrence in the accompanying parents (**f-k**) using BAC clone RP11-656O14 (red) which is located within the deleted region. The centromere 17 probe was included for reference (green). All individuals show only one signal for RP11-6565O14 (arrow indicating aberrant chromosome 17). The *de novo* occurrence for the deletion was proven in all individuals.

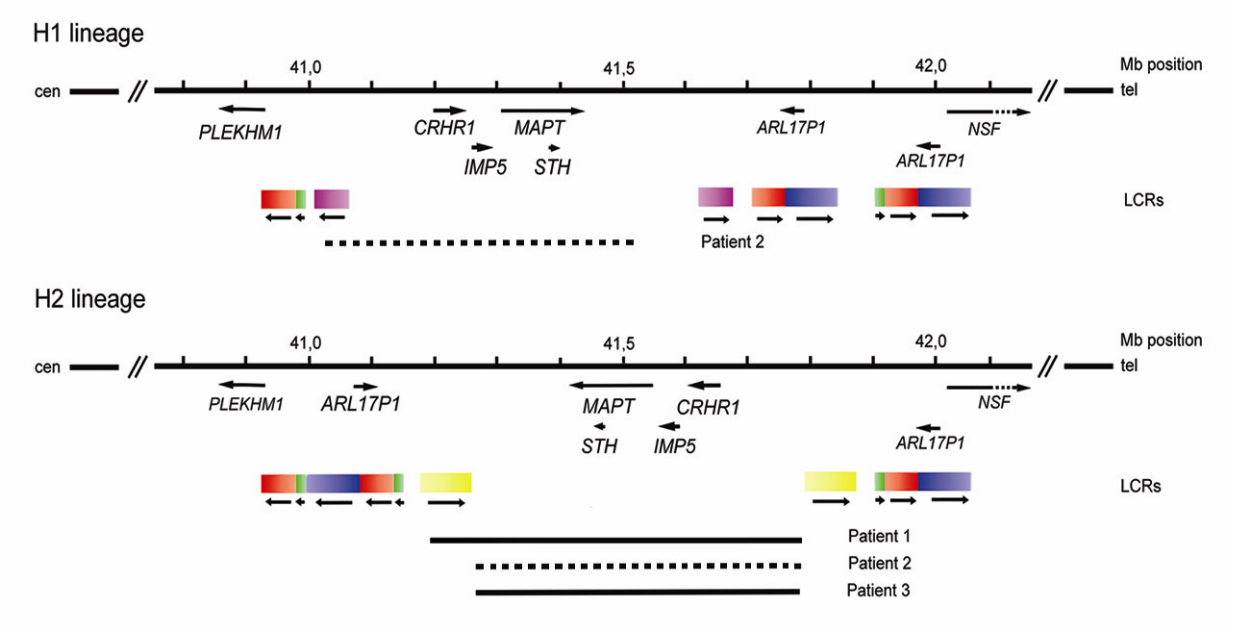


Figure 4.4
Transcript map and genomic architecture for the H1 and H2 lineage of the 17q21.31 region based on Cruts et al. (2005) and Takai et al. (2000). (Dotted) black lines represent the deleted 17q21.31 region identified in the three individuals. Genes located within this region are depicted by black arrows. The H1 and H2 lineage LCR17q structures are depicted as rectangles with colors signifying shared homology and horizontal arrows showing relative orientation (LCR17qA, red; LCR17qB, green; LCR17qC, purple; LCR17qD, blue; LCR17qE, yellow). Note the different genomic orientation of H2, the absence of LCR17qC and presence of LCR17qE. Cen, centromeric; Tel, telomeric.

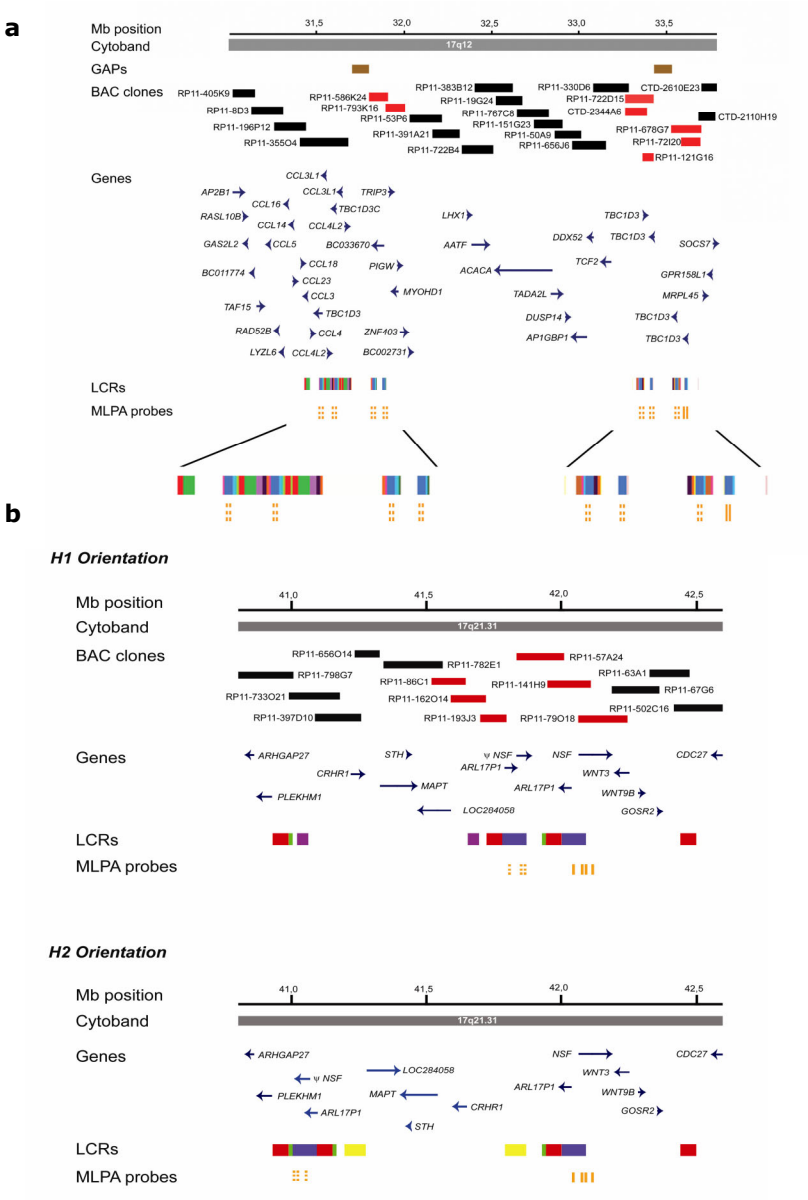


Figure 5.1
Examples of genome architecture and MLPA probe locations. (**a**), 17q12 (**b**) and 17q21.31. Note that for the 17q21.31, a 900-kb inversion polymorphism including the *NSF* locus has been observed in the general population, represented by the H1 and H2 orientation (Stefansson et al. 2005). For all regions, Mb position, cytoband, BAC clones, genes located within the region, LCRs/ GAPs (if present) and the location of the MLPA probes are depicted. The BAC clones that were previously reported to show CNV are colored red (Genome Variation Database). The LCR structures, if present, are depicted in colors to better represent their positional orientation with respect to each other. The locations of the MLPA probes are indicated by vertical orange solid lines, whereas dotted orange lines indicate alternative MLPA locations, sharing high homology to the region the probe was originally designed for.

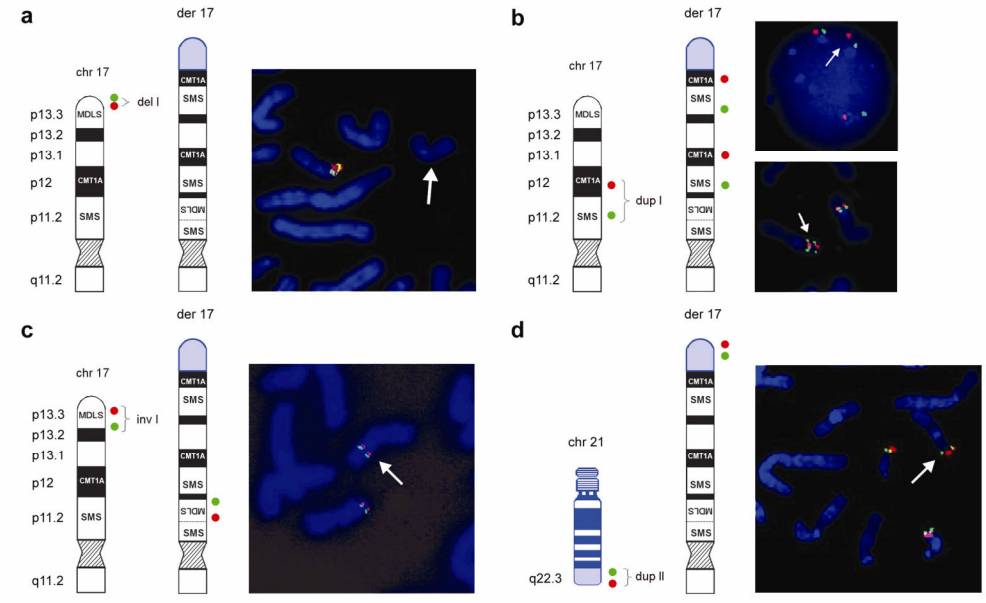


Figure 6.3
Schematic representation of a normal 17p and der(17) (black & white) with translocated chromosome 21 material (blue). The location of the FISH probes are shown on the left side of each figure panel; der(17) is indicated on FISH pictures by a white arrow. (**a**) Terminal deletion of 17pter was validated using BAC clones RP11-1260E13 (red) and CTD-2326F1 (green). (**b**) FISH with *PMP22*-specific PAC RP1-150M12 (red) and *RAI1*-specific BAC RP11-525O11 (green) revealed direct duplication of the CMT1A and SMS regions in 17p12p11.2. (**c**) FISH with PAC RP1-95H6 (red) and BAC GS-202L17 (green) showed inverted insertion of the MDLS region into the SMS region. (**d**) Array CGH also identified a duplication of 21q22.3. Additional FISH analysis using BAC clones RP11-40L10 (green) and RP11-16B19 (red) revealed that the duplicated material of 21q22.3 was translocated onto der(17). Summary of FISH results is provided in **Table 6.1**.

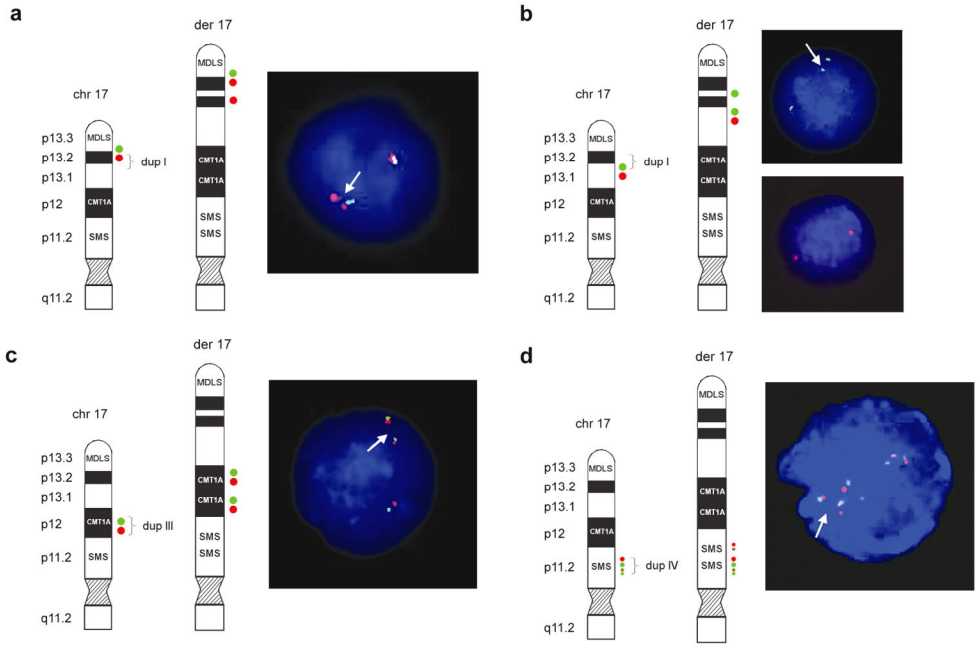


Figure 6.4
Schematic representation of the normal 17p and der(17). The locations of the FISH probes are shown on the left side of each figure panel. (**a**) Duplication I showed relatively simple fluorescence signal pattern with probes RP11-810M2 (green; normal) and RP11-597I9 (red; duplicated) for the distal breakpoint, and (**b**) RP11-222J21 (green; duplicated) and RP11-98D15 (red; normal) for the proximal breakpoint. (**c**) Direct orientation of duplication III was shown using BAC clones RP11-601N13 (green) and RP11-726O12 (red). (**d**) For the distal breakpoint of duplication IV, BAC clones RP11-448D22 (green) and CTD-2145A24 (red) showed duplicated signals on der(17), indicating that both middle SMS-REP and LCR17pB are duplicated as a block. Note four red signals on der(17) representing two normal and two duplicated copies of LCR17pA/B (dupIII) and LCR17pB (dupIV) and four green signals depicting three normal copies of SMS-REPs and the duplicated middle SMS-REP. Summary of FISH results is provided in **Table 6.2**.

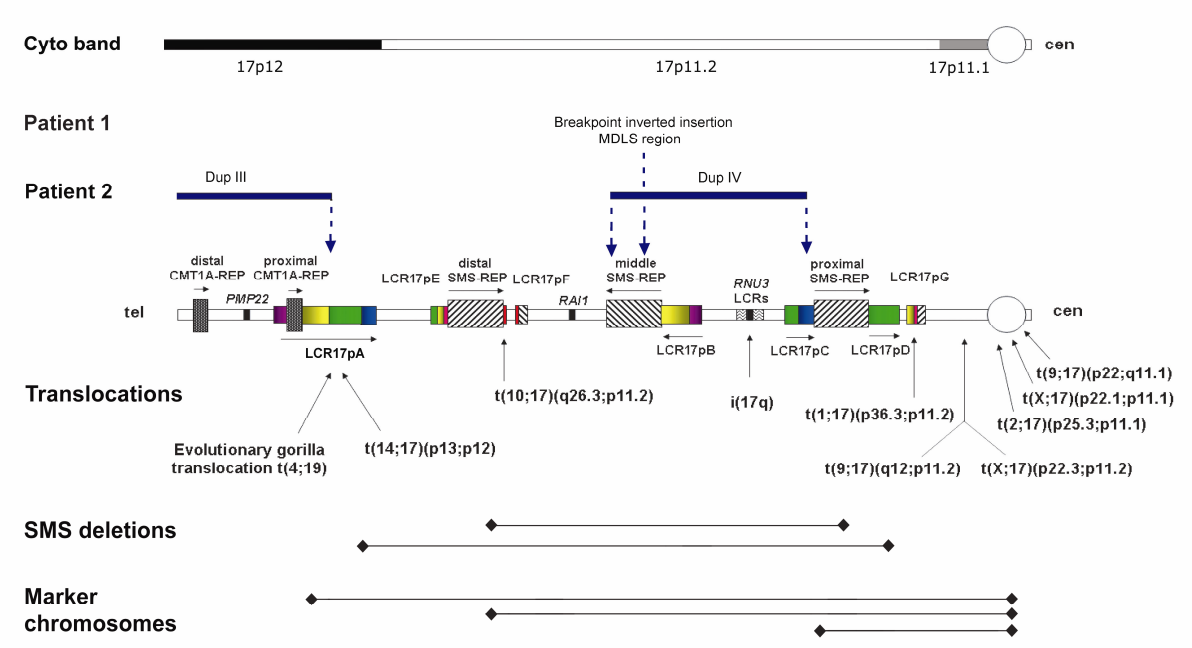


Figure 6.5

Top: Cytogenetic chromosome bands 17p12, and sub-bands 17p11.2 and 17p11.1 are shown. Breakpoints of the chromosome rearrangements in patients 1 and 2 associated with architectural features in proximal 17p are shown by vertical dotted blue arrows. In patient 1, one breakpoint is located within the middle SMS-REP/LCR17pB block, whereas in patient 2 breakpoints are located within LCR17pA (proximal Dup III), middle SMS-REP (distal Dup IV) and LCR17pC (proximal Dup IV). Duplication III and IV in patient 2 are indicated by horizontal blue lines. **Bottom**: Previously identified rearrangements associated with LCRs in 17p. Breakpoints of translocations and isochromosome 17q are indicated by vertical black arrows whereas common ~4 Mb and uncommon ~5 Mb SMS deletions and marker chromosomes are indicated by black horizontal lines. The LCR17p structures are depicted in colors to better represent their positional orientation with respect to each other; the shaded rectangles and horizontal black arrows represent the orientation of the LCRs.

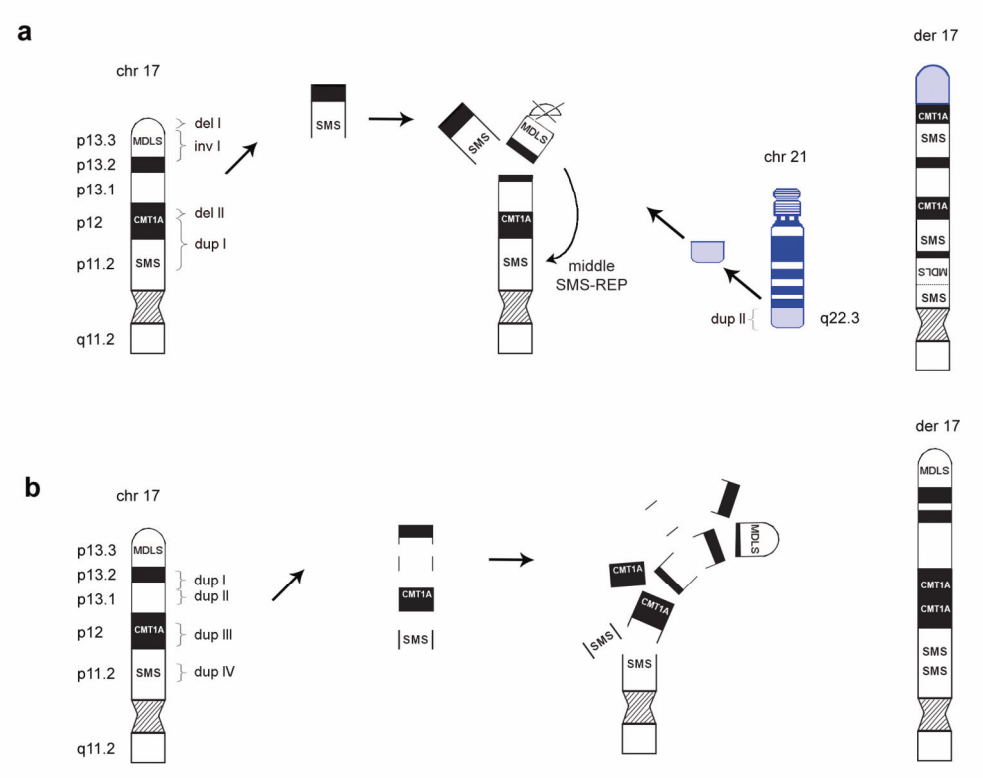


Figure 6.6
Schematic representation of the proposed mechanism for CCR formation in patient 1 and patient 2. (**a**) The complex rearrangement observed in patient 1 includes ten breakpoints, with one insertion, two microdeletions and two microduplications. We propose that the breaks in chromosome 17, p13.3, p12 and p11.2 as well as the break in chromosome 21q22.3 arose simultaneously. This resulted in the insertion of the MDLS region into the middle SMS-REP/LCR17pB block, loss of telomeric 17p and part of 17p12 and duplication of the CMT1A region. Additionally, the duplicated 21q22.3 fragment was translocated to the 17p subtelomeric region. (**b**) The complex rearrangement in patient 2 included eight breaks. We suggest that breaks in chromosome 17, p13.2, p13.1, p12 and p11.2 occurred at the same time, resulting in four interspersed directly orientated microduplications. Note: arrows do not represent a chronological order of events.